



**Etude moléculaire et structurale d'une intégrase
rétrovirale pour le développement de nouveaux
antirétroviraux et étude cristallographique d'
-galactosidases thermostables issues du microorganisme
Geobacillus stearothermophilus**

Romain Merceron

► **To cite this version:**

Romain Merceron. Etude moléculaire et structurale d'une intégrase rétrovirale pour le développement de nouveaux antirétroviraux et étude cristallographique d' -galactosidases thermostables issues du microorganisme Geobacillus stearothermophilus. Virologie. Université Claude Bernard - Lyon I, 2013. Français. NNT : 2013LYO10189 . tel-01146688

HAL Id: tel-01146688

<https://theses.hal.science/tel-01146688>

Submitted on 28 Apr 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

N° d'ordre : 189-2013

Année 2013

THESE DE L'UNIVERSITE DE LYON
Spécialité Biochimie

Délivrée par

L'UNIVERSITE CLAUDE BERNARD LYON 1

ECOLE DOCTORALE INTERDISCIPLINAIRE SCIENCE-SANTE

DIPLÔME DE DOCTORAT

(Arrêté du 7 août 2006)

soutenue publiquement le 4 Novembre 2013

par

Romain MERCERON

Etude moléculaire et structurale d'une intégrase rétrovirale
pour le développement de nouveaux antirétroviraux et étude
cristallographique d' α -galactosidases thermostables issues du
microorganisme *Geobacillus stearothermophilus*

Directeur de thèse : Monsieur le Professeur Patrice GOUET
Co-directrice de thèse : Madame le Docteur Corinne RONFORT

JURY :

Madame	le Docteur	S. MORÉRA	Rapporteur
Monsieur	le Docteur	V. PARISSI	Rapporteur
Monsieur	le Docteur	S. TRANIER	Examineur
Monsieur	le Professeur	R. BUCHET	Examineur
Monsieur	le Professeur	P. GOUET	Directeur de thèse
Madame	le Docteur	C. RONFORT	Co-directrice de thèse

UNIVERSITE CLAUDE BERNARD - LYON 1

Président de l'Université

Vice-président du Conseil d'Administration

Vice-président du Conseil des Etudes et de la Vie
Universitaire

Vice-président du Conseil Scientifique

Directeur Général des Services

M. François-Noël GILLY

M. le Professeur Hamda BEN HADID

M. le Professeur Philippe LALLE

M. le Professeur Germain GILLET

M. Alain HELLEU

COMPOSANTES SANTE

Faculté de Médecine Lyon Est – Claude Bernard

Faculté de Médecine et de Maïeutique Lyon Sud –
Charles Mérieux

Faculté d'Odontologie

Institut des Sciences Pharmaceutiques et
Biologiques

Institut des Sciences et Techniques de la
Réadaptation

Département de formation et Centre de Recherche
en Biologie Humaine

Directeur : M. le Professeur J. ETIENNE

Directeur : Mme la Professeure C. BURILLON

Directeur : M. le Professeur D. BOURGEOIS

Directeur : Mme la Professeure C.
VINCIGUERRA

Directeur : M. le Professeur Y. MATILLON

Directeur : M. le Professeur P. FARGE

COMPOSANTES ET DEPARTEMENTS DE SCIENCES ET TECHNOLOGIE

Faculté des Sciences et Technologies

Département Biologie

Département Chimie Biochimie

Département GEP

Département Informatique

Département Mathématiques

Département Mécanique

Département Physique

Département Sciences de la Terre

UFR Sciences et Techniques des Activités

Physiques et Sportives

Observatoire des Sciences de l'Univers de Lyon
Polytech Lyon

Ecole Supérieure de Chimie Physique Electronique

Institut Universitaire de Technologie de Lyon 1

Institut Universitaire de Formation des Maîtres

Institut de Science Financière et d'Assurances

Directeur : M. le Professeur F. DE MARCHI

Directeur : M. le Professeur F. FLEURY

Directeur : Mme le Professeur H. PARROT

Directeur : M. N. SIAUVE

Directeur : M. le Professeur S. AKKOUCHE

Directeur : M. le Professeur A. GOLDMAN

Directeur : M. le Professeur H. BEN HADID

Directeur : Mme S. FLECK

Directeur : Mme la Professeure I. DANIEL

Directeur : M. C. COLLIGNON

Directeur : M. B. GUIDERDONI

Directeur : M. P. FOURNIER

Directeur : M. G. PIGNAULT

Directeur : M. C. VITON

Directeur : M. A. MOUGNIOTTE

Administrateur provisoire : M. N. LEBOISNE

REMERCIEMENTS

Je souhaite tout d'abord remercier le Pr. Gilbert Deléage, directeur de l'Institut de Biologie et Chimie des Protéines, de m'avoir accueilli au sein de son unité de recherche.

J'adresse mes sincères remerciements au Dr Richard Haser qui m'a permis d'intégrer son équipe. Je voudrais ensuite remercier chaleureusement mon directeur de thèse, le Pr. Patrice Gouet. Je lui suis reconnaissant pour sa disponibilité, ses qualités pédagogiques et scientifiques, ainsi que sa bonne humeur. J'ai beaucoup appris à ses côtés et je le remercie pour tout cela. J'adresse également mes sincères remerciements à ma co-directrice de thèse, le Dr Corinne Ronfort, pour tous ses conseils avisés mais également sa disponibilité tout au long de ce travail.

J'adresse également toute ma gratitude au Dr Solange Moréra et au Dr Vincent Parissi pour avoir accepté d'être rapporteurs de cette thèse ainsi qu'au Dr Samuel Tranier et au Pr. René Buchet qui ont accepté d'examiner mes travaux.

Je tiens bien sûr à remercier l'ensemble des membres de l'équipe Rétrovirus, Intégration Virale et Vaccins du laboratoire Rétrovirus et Pathologie Comparée. Cela a été un véritable plaisir de travailler avec vous. Je remercie tout particulièrement Marie-Pierre (miss MTT), Kathy (miss Nobel Prize), Barbara (miss reptiles, je ne t'en veux pas...), Franck (mister droso) et bien sûr la célèbre miss « Momo ». J'ai vraiment apprécié travailler à vos côtés et je garde un très bon souvenir de mon séjour dans votre bureau (et de notre passage à Fréjus). Merci aussi à Saw-See et Pierre pour tous leurs conseils et leur sympathie. Je remercie également chaleureusement Karen, Alain, Sylvie, Catherine et d'autres que je pourrais oublier et je m'en excuse par avance.

Je tiens évidemment à associer à ces remerciements l'ensemble des membres de l'équipe de Biocristallographie. Ma première pensée va tout d'abord à Allison, ma chef spirituelle. Je te remercie pour la grande disponibilité dont tu as fait preuve pour moi, lorsque je suis arrivé dans l'équipe, ainsi que pour ta gentillesse. Je te remercie sincèrement. Merci bien sûr à Nushin et Patrice. Merci également à Michel, j'espère que tu passes une retraite paisible. Merci à mister Robert (prononciation à l'américaine) pour les horoscopes et nouvelles « pourries » et pour m'avoir sauvé la vie à de nombreuses reprises à cause d'un ordinateur capricieux. Merci à Christophe pour tous tes bons conseils scientifiques ainsi que pour ton assistance dans les attaques « canidés ». Merci à Sébastien, mon conseiller « avenir », pour toutes nos discussions sur « l'après thèse, mode d'emploi » et toutes tes bonnes astuces (cf ton système de dialyse). Merci à Lionel, pour nos craquages du matin, du midi et du soir. Merci au breton (Frédéric Abdel Bachir). Merci à Alex (peu cher), Jennifer (Romain tu arrêtes) pour ces années que nous avons passées ensemble au laboratoire. Merci à Magalie (et non Magali) pour tout ce que tu as pu m'apprendre sur la vie des canidés, en particulier grâce à ton site internet riche en informations, tu es vraiment une passionnée et c'est tant mieux (j'oubliais, et aussi pour m'avoir supporté). Merci à Preeti et Urvasi pour leur bonne humeur quotidienne. Merci aussi au petit nouveau, Loïc, ainsi qu'à Halima, je vous souhaite que de bonnes choses pour la suite.

Merci aux membres du plateau de cristallogénèse (Virginie, Marie-Pierre). Merci également aux membres du plateau technique de spectrométrie de masse (Frédéric, Isabelle), pour votre disponibilité et votre sympathie. Je remercie également sincèrement Roland de la plateforme de biophysique pour sa disponibilité et ses nombreux conseils.

Une pensée aussi pour les autres IBCPiens, Franck, Jérémy, Carine, Vincent, Thomas et Burcu. Je vous remercie tous pour ces années agréables passées à l'IBCP. Une pensée également pour les non IBCPiens Julie, Lorraine, Lama et également Aubin.

Une pensée également à Liçou, Célinette, Max, Nerixm, Kiki, Patoch, Pierro, Totof, Pakal, Lainlain, Laurent, Simon et JB.

Un clin d'œil à messieurs Turpin et De Rossi, sinon ils m'en voudraient.

Enfin un grand merci à ma famille en particulier à maman, papa et Michel, le frère et ma belle soeur Martine et mes petites Léa, Justine et Rose.

Je terminerai en remerciant Marie, pour ta confiance et ton soutien sans faille pendant ces années.

RÉSUMÉ

L'intégrase (IN) est une protéine clé du cycle de réplication des rétrovirus et constitue une cible thérapeutique importante pour le développement d'antirétroviraux. Cette enzyme intervient dans la phase précoce du cycle réplcatif rétroviral, en catalysant l'intégration du génome viral rétro-transcrit dans celui de la cellule hôte. Les équipes de BioCristallographie et de Rétrovirologie du site de Lyon-Gerland ont découvert par cristallographie aux rayons X, une nouvelle possibilité d'assemblage dimérique du domaine central catalytique (CCD) de l'intégrase du *Rous Associated virus type 1* (RAV-1 IN). Dans le cadre de mon travail de thèse, un protocole de surproduction et de purification d'un mutant du domaine catalytique isolé (H103C) a été optimisé, afin de démontrer l'existence de cet assemblage en solution grâce à un pont disulfure inter-moléculaire. Différentes méthodes biochimiques et biophysiques ont été mises au point, afin de tester la capacité de petites molécules d'intérêt à se lier et à stabiliser ce « nouvel » assemblage. Un protocole de surproduction et de purification de l'IN entière du RAV-1 a également été développé et mis au point. Des essais de cristallisation et une étude par diffusion des rayons X aux petits angles (SAXS) ont été réalisés. Un mutant H103C de la protéine entière a été produit, afin de vérifier la formation de la « nouvelle » interface sur la protéine entière.

Le microorganisme *Geobacillus stearothermophilus* produit deux α -galactosidases thermostables, AgaA et AgaB, qui appartiennent à la famille GH36 des glycoside hydrolases. Ces deux isoenzymes partagent 97 % d'identité de séquence, mais ont des activités catalytiques différentes. Une collaboration a été initiée avec l'Institut de Génétique Industrielle de l'Université de Stuttgart (Allemagne), afin de mieux comprendre les spécificités catalytiques de ces deux isoenzymes. Les structures cristallines d'AgaA et AgaB ont été résolues ainsi que la structure du mutant AgaA A355E, qui présente des caractéristiques enzymatiques similaires à AgaB. L'analyse de ces trois structures montre que la substitution A355E, qui est située à 15 Å du site actif, entraîne un déplacement significatif du tryptophane W336 du sous-site catalytique -1. Ce mouvement peut expliquer les spécificités catalytiques des deux isoenzymes. Les structures cristallines d'AgaA A355E en complexe avec un inhibiteur compétitif (1-désoxygalactonojirimycine) et deux substrats (D-raffinose et D-stachyose) ont été obtenues. Elles montrent que la spécificité de reconnaissance du substrat se joue essentiellement au niveau du sous-site -1.

Summary

Integrase (IN) is a key protein in the retrovirus life cycle and constitutes an important therapeutic target for the development of antiretroviral compounds. This enzyme is involved in the early phase of the retroviral replication cycle and catalyses the retrotranscribed viral DNA integration into the host cell genome. The teams of BioCrystallography and Retrovirology of Lyon-Gerland demonstrated by X-ray crystallography, the existence of a new dimeric assembly of the central catalytic domain (CCD) of Rous Associated Virus type 1 integrase or RAV-1 IN. As part of my thesis work, a protocol of overproduction and purification of the H103C isolated catalytic domain mutant was developed to demonstrate the existence of this dimeric assembly in solution stabilized by an inter-molecular disulfide bond. Biochemical and biophysical methods were developed to test the ability of small molecules of interest to bind and stabilize this "new" assembly. A protocol of overproduction and purification of full length RAV-1 integrase was developed. Crystallization trials and SAXS studies were undertaken. The H103C mutant of the entire protein was produced to verify the formation of the "new" interface on the full length protein.

The microorganism *Geobacillus stearothermophilus* produces two thermostable α -galactosidases named AgaA and AgaB, which belong to the GH36 glycoside hydrolase family. These two isoenzymes share 97 % sequence identity, but have different catalytic properties. A collaborative study was initiated with the Institute of Industrial Genetics, University of Stuttgart (Germany), to better understand the catalytic specificity of these two isoenzymes. The crystal structures of AgaA and AgaB were solved in two different crystal systems. The crystal structure of the mutant AgaA A355E, which has catalytic properties similar to those of AgaB, was also determined. These three structures show that the A355E substitution, which is located 15 Å away from the active site, results in a significant displacement of the W336 tryptophan residue from the catalytic subsite -1. This could explain the catalytic specificities of the two isoenzymes. The crystal structures of AgaA A355E with a competitive inhibitor (1-deoxygalactonojirimycin) and with two substrates (D-raffinose and D-stachyose) were obtained. They show that subsite -1 is mainly responsible for the specificity of substrate recognition.

TABLE DES MATIÈRES

REMERCIEMENTS	5
RÉSUMÉ	7
TABLE DES MATIÈRES	9
ABREVIATIONS	15
CURRICULUM VITAE	19
AVANT-PROPOS	23

PARTIE 1 : ETUDE MOLÉCULAIRE ET STRUCTURALE D'UNE INTÉGRASE RÉTROVIRALE POUR LE DÉVELOPPEMENT DE NOUVEAUX ANTIRÉTROVIRAUX..... 25

EXPOSÉ BIBLIOGRAPHIQUE	27
<i>I. Les rétrovirus</i>	29
1. Historique	29
2. Taxonomie des rétrovirus.....	30
3. Organisation structurale et génomique	32
3.1. La particule virale	32
3.2. Le génome viral	33
3.2.1. Généralités	33
3.2.2. Structure.....	33
3.2.3. L'ARN génomique et les ARN messagers viraux.....	34
3.3. Les protéines virales enzymatiques et structurales	35
3.3.1. Cas des rétrovirus complexes.....	35
3.3.2. Cas des oncovirus	36
4. Assemblage et réplication des particules virales	37
4.1. Phase pré-intégrative	37
4.2. Phase post-intégrative.....	38
5. Pathogénèses rétrovirales.....	39
5.1. Les rétrovirus dans l'oncogénèse	39
5.2. Les rétrovirus dans l'immunodéficience.....	40
6. Restriction intracellulaire de l'infection	41
7. SIDA et stratégies antirétrovirales.....	41
<i>II. L'intégrase</i>	43
1. Intégration et réplication	43
2. Historique	44
3. Propriétés fonctionnelles	44
3.1. Intégrase et transposases.....	44
3.2. Fonctions enzymatiques principales	45
3.3. Fonctions enzymatiques secondaires.....	46
3.4. Fonctions non enzymatiques.....	47
4. Le complexe de pré-intégration	47

5. Propriétés structurales	49
5.1. Structures primaires	49
5.2. Structures tridimensionnelles	50
5.2.1. Le domaine amino-terminal ou NTD	50
5.2.2. Le domaine central catalytique ou CCD	51
5.2.3. Le domaine carboxy-terminal ou CTD	52
6. Etat oligomérique de la protéine entière	53
7. Les stratégies anti-intégrase	55
OBJECTIFS DE L'ÉTUDE	61
MATÉRIEL ET MÉTHODES	67
<i>I. Techniques préparatives</i>	<i>69</i>
1. Obtention et stockage des ADN	69
1.1. Souches bactériennes utilisées	69
1.2. Extraction des ADN plasmidiques par lyse alcaline	69
1.3. Purification des ADN	70
2. Manipulations enzymatiques des ADN	71
2.1. Ligature de fragments d'ADN	71
2.2. Digestion des ADN par les enzymes de restriction	71
2.3. Traitement des ADN par l'enzyme Dpn1	71
2.4. Amplification in vitro de l'ADN par PCR	72
2.5. Mutagenèse dirigée par PCR	72
3. Introduction des ADN dans les bactéries	73
4. Antibiotiques utilisés	74
5. Surproduction des protéines recombinantes	74
5.1. Souches bactériennes utilisées	74
5.2. Vecteurs plasmidiques utilisés	77
5.3. Méthodes de culture et surproduction des protéines recombinantes	79
6. Purification des protéines recombinantes	80
6.1. Extraction des protéines	80
6.2. Purification par chromatographie	81
6.2.1. Chromatographie d'affinité pour le métal	81
6.2.2. Chromatographie d'exclusion de taille	82
7. Traitement des protéines purifiées	83
7.1. Traitements enzymatiques	83
7.1.1. Clivage par la thrombine	83
7.1.2. Clivage par la protéase du TEV	83
7.2. Concentration des protéines	83
7.3. Echanges de tampons	84
<i>II. Techniques analytiques</i>	<i>84</i>
1. Contrôle de la qualité et de la quantité des ADN	84
1.1. Spectroscopie ultraviolet-visible	84
1.2. Electrophorèse sur gel d'agarose	85
1.3. Séquençage de l'ADN	86
2. Contrôle de la qualité et de la quantité des protéines	87
2.1. Analyses spectroscopiques	87

2.1.1. Spectroscopie ultraviolet-visible	87
2.1.2. Spectroscopie à corrélation de photons	88
2.1.3. Spectrométrie de masse.....	89
2.2. Analyses par électrophorèse sur gel d'acrylamide.....	89
2.2.1. En conditions dénaturantes	89
2.2.2. En conditions natives.....	91
2.2.3. Méthodes de révélations.....	92
2.2.3.A. Bleu de coomassie colloïdal	92
2.2.3.B. Nitrate d'argent.....	92
2.2.3.C. Immuno-empreinte	92
2.3. Analyse par pontage chimique des protéines	93
2.3.1. Pontage covalent au BisSulfoSuccinimidyl Suberate (BS3)	93
2.3.2. Pontage au glutaraldéhyde (GTA)	93
3. Mesure des interactions protéines/ligands.....	94
3.1. Analyse par Titration Calorimétrique Isotherme (ITC)	94
3.2. Analyse par mesure de la fluorescence intrinsèque des protéines	94
3.3. Analyse par fluorimétrie différentielle à balayage thermique.....	96
4. Analyses structurales.....	98
4.1. Analyse par diffusion des rayons X aux petits angles.....	98
4.2. Analyse par cristallographie aux rayons X.....	100
4.2.1. La cristallogénèse et les méthodes associées	100
4.2.1.A. Principe général	100
4.2.1.B. La diffusion en phase vapeur	101
4.2.1.C. Techniques de trempage et de co-cristallisation	101
4.2.1.D. Cryoprotection des cristaux	102
4.2.1.E. Production des rayons X et synchrotron	102
4.2.2. Obtention et enregistrement des données de diffraction	103
4.2.2.A. Principe de diffraction.....	103
4.2.2.B. Traitement des données	104
4.2.3. Le problème des phases.....	105
4.2.3.A. La méthode du remplacement isomorphe multiple	105
4.2.3.B. La méthode de dispersion anormale multiple.....	105
4.2.3.C. La méthode du remplacement moléculaire	105
4.2.3.C.α. Fonction de rotation	106
4.2.3.C.β. Fonction de translation	106
4.2.4. Affinement des modèles	107
4.2.5. Les cartes de densité électronique.....	108
RÉSULTATS ET DISCUSSION	109
<i>I. Etude moléculaire et structurale du domaine catalytique de l'intégrase du RAV-1.....</i>	<i>111</i>
1. Intégrase IN-CCD-H103C.....	111
1.1. Modélisation moléculaire.....	111
1.2. Surproduction et purification	112
1.3. Caractérisation de la protéine en solution.....	113
1.4. Optimisation de la surproduction	115
1.5. Formation du pont disulfure <i>in vitro</i>	116

1.6. Essais de cristallisation	118
1.7. Interactions de petites molécules avec le nouvel assemblage	119
1.7.1. Conception des molécules.....	119
1.7.2. Solubilisation des composés.....	121
1.7.3. Mesure <i>in vitro</i> des interactions entre le nouvel assemblage et les composés synthétisés.....	122
1.7.3.A. Mesure par calorimétrie de titration isotherme ou ITC.....	122
1.7.3.B. Mesure par atténuation de fluorescence intrinsèque des protéines	123
1.7.3.C. Mesure par fluorimétrie différentielle à balayage thermique	124
2. Intégrase IN-CCD-WT	127
2.1. Effets des composés sur l'état d'oligomérisation d' IN-CCD-WT	127
2.2. Trempages des cristaux d'IN-CCD-WT et co-cristallisations avec les composés	129
2.3. Trempages des cristaux d'IN-CCD-WT et co-cristallisations avec des nucléotides	130
3. Conclusion	132
<i>II. Etude moléculaire et structurale de l'intégrase entière du RAV-1</i>	<i>133</i>
1. Intégrase INFLWT-pET-30a	133
1.1. Surproduction et solubilisation	133
1.2. Purification	134
1.3. Caractérisation de l'échantillon.....	136
1.3.1. Analyse par spectrométrie de masse	136
1.3.2. Analyse par spectroscopie à corrélation de photons	137
1.3.3. Analyse par chromatographie d'exclusion de taille	138
1.3.4. Analyse par pontage covalent au Bis SulfoSuccinimidyl Suberate (BS3).....	139
1.4. Essais de cristallisation	140
1.5. Diffusion des rayons X aux petits angles (SAXS).....	141
1.6. Conclusion	143
2. Intégrase INFLWT-pPROEX-HTa.....	144
2.1. Surproduction et solubilisation	144
2.2. Essais de solubilisation	146
2.3. Purification	150
2.4. Clivage de l'étiquette de purification	152
2.5. Caractérisation de l'échantillon.....	154
2.5.1. Analyse par pontage covalent au Bis SulfoSuccinimidyl Suberate (BS3) et glutaraldéhyde	154
2.5.2. Analyse par spectroscopie à corrélation de photons	155
2.6. Essais de cristallisation	155
2.7. Conclusion	155
3. Intégrase INFLH103C-pPROEX-HTa	156
3.1. Purification et caractérisation de l'échantillon	156
3.2. Conclusion	158
CONCLUSION ET PERSPECTIVES	159
 PUBLICATION 1 : A CRYSTAL STRUCTURE OF THE CATALYTIC CORE DOMAIN OF AN AVIAN SARCOMA AND LEUKEMIA VIRUS INTEGRASE SUGGESTS AN ALTERNATE DIMERIC ASSEMBLY.....	167

PARTIE 2 : ETUDE CRISTALLOGRAPHIQUE D'ALPHA-GALACTOSIDASES THERMOSTABLES ISSUES DU MICROORGANISME *GEOBACILLUS STEAROTHERMOPHILUS*..... 179

EXPOSÉ BIBLIOGRAPHIQUE 181

I. Les glycoside hydrolases 183

1. Généralités 183

2. Classification 184

3. Mécanismes catalytiques 184

3.1. Hydrolyse par le mécanisme de rétention 186

3.2. Hydrolyse par le mécanisme d'inversion..... 188

3.3. Autres mécanismes 189

4. Glycoside hydrolases et synthèse..... 189

5. Topologie des sites actifs..... 189

II. Les Alpha-galactosidases..... 190

1. Généralités 190

2. Classification 191

3. Sources et rôles physiologiques 192

4. Applications 192

5. Propriétés structurales 193

III. Les extrémozymes..... 196

*IV. Alpha-galactosidases de *Geobacillus stearothermophilus*..... 197*

1. Généralités 197

2. Propriétés catalytiques..... 198

3. Applications 198

OBJECTIFS DE L'ÉTUDE..... 201

MATÉRIEL ET MÉTHODES 205

1. Etude structurale 207

1.1. Surproduction et purification des protéines 207

1.2. Essais de cristallisation 207

1.3. Trempages et cryoprotection 208

RÉSULTATS ET DISCUSSION 209

1. Résultats obtenus 211

2. Résumé de l'article et de l'analyse des structures 212

CONCLUSION ET PERSPECTIVES 217

PUBLICATION 2 : THE MOLECULAR MECHANISM OF THERMOSTABLE ALPHA-GALACTOSIDASES AGAA AND AGAB EXPLAINED BY X-RAY CRYSTALLOGRAPHY AND MUTATIONAL STUDIES 221

RÉFÉRENCES BIBLIOGRAPHIQUES 234

ABREVIATIONS

Å	Ångström (1 Å = 10 ⁻¹⁰ mètre)
A	Absorbance
ADN	Acide DésoxyriboNucléique
ALV	<i>Avian Leukemia Virus</i>
AMP-CPP	Adenosine-5'-[(α,β)-methyleno]triphosphate
AMP-PNP	Adenosine-5'-[(β,γ)-imido]triphosphate
AmSO ₄	Sulfate d'ammonium
ARN	Acide RiboNucléique
ARNm	ARN messenger
ARNt	ARN de transfert
ASLV	<i>Avian Sarcoma and Leukemia Virus</i>
ASV	<i>Avian Sarcoma Virus</i>
ATP	Adénosine Tri-Phosphate
Att	<i>Attachment</i>
AZT	AZidoThymidine
BCIP-NBT	5-Bromo-4-Chloro-3-Indolyl Phosphate (BCIP)-NitroBlue Tetrazolium (NBT)
BET	Bromure d'ETHidium
BFV	<i>Bovine Foamy Virus</i>
BLV	<i>Bovine Leukemia Virus</i>
BS3	<i>Bis (SulfoSuccinimidyl) Suberate</i>
°C	Degré Celsius
C-	Carboxy
CA	Protéine de CApside
CAZy	<i>Carbohydrate-Active enZymes</i>
CCD	<i>Catalytic Core Domain</i>
CD	Cluster de Différenciation
CTD	<i>C-Terminal Domain</i>
Da	Dalton
ddNTP	Di désoxyriboNucléotide TriPhosphate
dNTP	DésoxyriboNucléotide TriPhosphate
DLS	<i>Dynamic Light Scattering</i>
DMSO	DiMethyl SulfOxide
DTT	DiThioThréitol
EC	<i>Enzyme Commission</i>
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	<i>EthyleneDiamineTetraacetic Acid</i>
EIAV	<i>Equine Infectious Anemia Virus</i>
env	<i>Envelope</i>
ESRF	<i>European Synchrotron Radiation Facility</i>
FDA	<i>Food and Drug Administration</i>
FELV	<i>Feline Leukemia Virus</i>
FIP	<i>French beamline for Investigation of Proteins</i>
FIV	<i>Feline Immunodeficiency Virus</i>
FL	<i>Full Length</i>
FPLC	<i>Fast Protein Liquid Chromatography</i>
gag	<i>Group-specific antigen</i>
GALV	<i>Gibbon Ape Leukemia Virus</i>
GH	Glycoside Hydrolase

GTA	GluTAraldéhyde
HAART	<i>Highly Active AntiRetroviral Therapy</i>
HCl	Acide Chlorhydrique
HEPES	<i>N-2-HydroxyEthylPiperazine-N'-2-EthaneSulfonic acid</i>
HERV	<i>Human Endogenous RetroVirus</i>
HFV	<i>Human Foamy Virus</i>
HTLV	<i>Human T-lymphotropic virus</i>
ICTV	<i>International Committee on Taxonomy of Viruses</i>
ID	<i>Insertion Device</i>
IN	INtégrase
IPTG	IsoPropylThioGalactoside
IUBMB	<i>International Union of Biochemistry and Molecular Biology</i>
kb	Kilobase
Kcat	Constante catalytique
Kd	Constante de dissociation
Ki	Constante d'inhibition
kDa	kiloDalton
Km	Constante de Michaelis-Menten
LAV	<i>Lymphadenopathy Associated Virus</i>
LB	<i>Lysogeny Broth</i>
LEDGF	<i>Lens Epithelium-Derived Growth Factor</i>
LTR	<i>Long Terminal Repeat</i>
MA	Protéine de MATrice
MALDI-TOF	<i>Matrix-Assisted Laser Desorption/Ionization Time-Of-Flight</i>
MES	<i>2-(N-morpholino)EthaneSulfonic acid</i>
Min	Minute
MLV	<i>Murine Leukemia Virus</i>
MM	Masse Moléculaire
MMTV	<i>Mouse Mammary Tumor Virus</i>
MPMV	<i>Mason Pfizer Monkey Virus</i>
MULV	<i>Murine Leukemia Virus</i>
N-	Amino
NaCl	Chlorure de sodium
NATA	<i>N-Acetyl-L-TryptophAnamide</i>
NC	NucléoCapside
NHS	<i>N-HydroxySulfosuccinimide</i>
Ni-IDA	<i>Nickel - IminoDiacetic Acid</i>
Ni-NTA	<i>Nickel - NitriloTriacetic Acid</i>
nm	Nanomètre
NTD	<i>N-Terminal Domain</i>
PAGE	<i>PolyAcrylamide Gel Electrophoresis</i>
pb	Paire de bases
PCR	<i>Polymerase Chain Reaction</i>
PDB	<i>Protein Data Bank</i>
PEG	PolyEthylène Glycol
PES	PolyEtherSulfone
PFV	<i>Prototype Foamy Virus</i>
PHV	<i>Perch Hyperplasia Virus</i>
PIC	<i>Pre-Integration Complex</i>
pol	<i>Polymerase</i>
PPT	<i>PolyPurine Tract</i>
PR	PRotéase

RAV-1	<i>Rous Associated Virus type 1</i>
Rg	Rayon de giration
RMN	Résonance Magnétique Nucléaire
RMSD	<i>Root Mean Square Deviation</i>
RSV	<i>Rous Sarcoma Virus</i>
RT	<i>Reverse Transcriptase</i>
RTC	<i>Reverse Transcription Complex</i>
SAXS	<i>Small Angle X-ray Scattering</i>
SDS	<i>Sodium Dodecyl Sulfate</i>
SIDA	Syndrome d'ImmunoDéficiency Acquis
SIV	<i>Simian Immunodeficiency Virus</i>
SLS	Source de Lumière Suisse
SnRV	<i>Snakehead RetroVirus</i>
SU	Glycoprotéine Env de SURface
TAE	Tampon Tris, Acétate, EDTA
TEMED	N, N, N', N'-tétraméthyléthylènediamine
TEV	<i>Tobacco Etch Virus</i>
TF	Transformée de Fourier
TIM	<i>Triose-phosphate IsoMerase</i>
TM	Glycoprotéine TransMembranaire
Tm	<i>Melting Temperature</i>
TNT	Tampon Tris-NaCl-Tween
Tris	Tris-(hydroxyméthyl)-aminométhane
VIH	Virus de l'Immunodéficiency Humaine
Vm	Vitesse maximale
WDSV	<i>Walleye Dermal Sarcoma Virus</i>
WEHV	<i>Walleye Epidermal Hyperplasia Virus</i>
WT	<i>Wild Type</i>
(V/V)	Volume/volume
(W/V)	Poids/volume
(W/W)	Poids/poids

CURRICULUM VITAE

MERCERON Romain



Sexe : masculin

Date de naissance : 23 septembre 1986

Lieu de naissance : Cholet (Maine et Loire)

Nationalité : Française

Coordonnées professionnelles :

Institut de Biologie et Chimie des Protéines (IBCP)-CNRS FR 3302

Bases Moléculaires et Structurales des Systèmes Infectieux (BMSSI)-CNRS UMR 5086-Université de Lyon-UMS 3444 BioSciences Gerland-Lyon Sud

Equipe Biocristallographie et Biologie Structurale des Cibles Thérapeutiques (BBSCT)

7, Passage du Vercors-69367 LYON cedex 07

Tél. : +33 (0)4 72 72 26 31

Fax : +33 (0)4 72 72 26 16

Courriel : romain.merceron@ibcp.fr

CURSUS

- **2010-2013 : Doctorat de biochimie structurale et fonctionnelle**

Université Claude Bernard, LYON 1 (Rhônes-Alpes)

Bourse ministérielle EDISS

Sujet de thèse : « Etude moléculaire et structurale d'une intégrase rétrovirale pour le développement de nouveaux antirétroviraux et étude cristallographique d' α -galactosidases thermostables issues du microorganisme *Geobacillus stearothermophilus* »

Directeur de thèse : Pr. Patrice GOUET

Co-directrice de thèse : Dr Corinne RONFORT

Laboratoire d'accueil : Institut de Biologie et Chimie des Protéines (IBCP)

Unité Mixte de Recherche 5086 – CNRS-UCB LYON 1

Laboratoire Bases Moléculaires et Structurales des Systèmes Infectieux (BMSSI)

Equipe Biocristallographie et Biologie Structurale des Cibles Thérapeutiques (BBSCT)

- **2009-2010 : Master 2 Recherche Biochimie Structurale et Fonctionnelle**

Université Claude Bernard, LYON 1 (Rhônes-Alpes)

Stage et sujet de mémoire : « Etude structurale de l'intégrase rétrovirale aviaire » (10 mois)

Directeur de stage : Pr. Patrice GOUET

Laboratoire d'accueil : Institut de Biologie et Chimie des Protéines (IBCP)

Unité Mixte de Recherche 5086 – CNRS-UCB LYON 1

Laboratoire Bases Moléculaires et Structurales des Systèmes Infectieux (BMSSI)

Equipe Biocristallographie et Biologie Structurale des Cibles Thérapeutiques (BBSCT)

Collaboration : Dr Corinne RONFORT

Unité Mixte de Recherche 754 INRA-ENVL-EPHE-UCB LYON 1

Equipe Rétrovirus, Intégration Virale et Vaccins (RI2V)

Laboratoire Rétrovirus et Pathologie Comparée (RPC)

- **2008-2009 : Master 1 Biochimie**

Université Claude Bernard, LYON 1 (Rhônes-Alpes)

- **2006-2008 : Licence Biologie Biochimie spécialité Biologie Moléculaire option Bioinformatique Structurale**

Faculté des Sciences et des Techniques, Université de NANTES (Loire-Atlantique)

Stage volontaire : « Ingénierie de nouvelles activités transglycosidases et glycosynthases par évolution dirigée » (2 mois)

Directeur de stage : Pr. Jean-Pierre SINE

Laboratoire d'accueil : unité de Biotechnologie, Biocatalyse et Biorégulation (U3B)

Unité Mixte de Recherche 6204 - CNRS - Université de NANTES

- **2004-2006 : Première année de médecine (PCEM1)**

Faculté de Médecine, Université de NANTES (Loire-Atlantique)

Stage d'entrée en seconde année (PCEM2) : stage infirmier (2 mois)

Lieu : CHU Hôtel-Dieu, Bloc opératoire des urgences, NANTES (Loire-Atlantique)

Réalisation : soins infirmiers, accompagnement des patients, découverte de la pratique médicale en milieu hospitalier

- **2004 : Baccalauréat scientifique spécialité science de la vie et de la terre**

Lycée Saint-Gabriel, Saint-Laurent sur Sèvre (Vendée)

PUBLICATIONS

1. **Merceron R.**, Foucault M., Haser R., Mattes R., Watzlawick H., Gouet P. (2012) The molecular mechanism of thermostable alpha-galactosidases AgaA and AgaB explained by x-ray crystallography and mutational studies. *J Biol Chem* 287: 39642-39652.

2. Ballandras A, Moreau K, Robert X, Confort MP, **Merceron R**, Haser R., Ronfort C., Gouet P. (2011) A crystal structure of the catalytic core domain of an avian sarcoma and leukemia virus integrase suggests an alternate dimeric assembly. *PLoS One* 6: e23032.

PARTICIPATION A DES COMMUNICATIONS ORALES OU POSTERS

1. 16^{ème} journée scientifique du GLGS (Groupe Lyonnais de GlycoSciences)

Université Claude Bernard - Lyon 1 - France - 15 novembre 2012

Communication orale : « Le mécanisme moléculaire des α -galactosidases AgaA et AgaB de la famille GH36 expliqué par cristallographie aux rayons X » (Romain Merceron, Marine Foucault, Richard Haser, Ralf Mattes, Hildegard Watzlawick, Patrice Gouet)

2. 17^{ème} journée scientifique de l'EDISS (Ecole Doctorale Inter-disciplinaire Sciences-Santé)

Université Claude Bernard - Lyon 1 - France - 18 octobre 2012

Communication par affiche : « Etudes moléculaires et structurales d'intégrases rétrovirales pour le développement d'antirétroviraux » (Romain Merceron, Halima Yajjou, Allison Ballandras, Xavier Robert, Karen Moreau, Marie-Pierre Confort, Kathy Gallay, Corinne Ronfort, Patrice Gouet)

Affiche récompensée par le **prix des meilleurs posters** de la journée scientifiques de l'EDISS

3. Colloque français de biologie structurale (GTBio)

Montpellier - France - 3-6 avril 2012

Communication orale : « Le mécanisme moléculaire des α -galactosidases AgaA et AgaB de la famille GH36 expliqué par cristallographie aux rayons X » (Romain Merceron, Marine Foucault, Richard Haser, Ralf Mattes, Hildegard Watzlawick, Patrice Gouet)

4. 4^{ème} Meeting international sur l'intégration rétrovirale

Sienne - Italie - 4-7 octobre 2011

Communication par affiche : « Structural studies of an Avian Leukosis Virus Integrase » (Romain Merceron, Allison Ballandras, Xavier Robert, Cécile Méricoux, Patrice Vachette, Karen Moreau, Marie-Pierre Confort, Kathy Gallay, Saw-See Hong, Corinne Ronfort, Patrice Gouet)

5. Journées d'Animation Scientifique du Département de Santé Animale (JAS) de l'INRA

Fréjus - France - 22-25 mai 2011

Communication par affiche : « Etude structurale de l'intégrase d'un virus de la leucose aviaire » (Romain Merceron, Allison Ballandras, Xavier Robert, Cécile Méricoux, Patrice Vachette, Karen Moreau, Marie-Pierre Confort, Kathy Gallay, Saw-See Hong, Corinne Ronfort, Patrice Gouet)

AVANT-PROPOS

Ce travail de thèse a débuté en octobre 2010. Il a été réalisé sous la direction du Pr. Patrice Gouet au sein de l'équipe de « Biocristallographie et Biologie Structurale des Cibles thérapeutiques » du laboratoire « Bases Moléculaires et Structurales des Systèmes Infectieux » (BMSSI-IBCP, UMR5086 CNRS Université Lyon) et sous la co-direction du Dr Corinne Ronfort dans l'équipe Rétrovirus, Intégration Virale et Vaccins (RI2V) du laboratoire Rétrovirus et Pathologie Comparée (INRA, Lyon, UMR 754 INRA-ENVL-UCBL-EPHE).

Ce mémoire présente les études moléculaires et structurales que j'ai réalisées au cours de ma thèse sur deux types de protéine : l'intégrase du rétrovirus RAV-1 (pour *Rous Associated Virus type 1*) et les α -galactosidases thermostables AgaA et AgaB de *Geobacillus stearothermophilus*.

La première partie de ce manuscrit concerne ainsi l'étude moléculaire et structurale de l'intégrase du rétrovirus RAV-1. Ce sujet, issu de la collaboration entre le Pr. Patrice Gouet et le Dr Corinne Ronfort, a fait l'objet d'une thèse réalisée et soutenue par le Dr A. Ballandras en 2010. Ses travaux ont permis de résoudre, par remplacement moléculaire, la structure du domaine catalytique central isolé (CCD) de l'intégrase du RAV-1. L'examen de cette structure a révélé que ce domaine pouvait s'assembler suivant une nouvelle structure quaternaire caractérisée par une interface dimérique à trois paires d'hélices α contre deux paires d'hélices α dans tous les assemblages dimériques « canoniques » de CCD d'intégrases observés jusqu'à présent. L'hypothèse avancée est que cette nouvelle association dimérique, qui semble significativement différente de celle observée pour la phase d'intégration, pourrait se produire sur l'enzyme entière et engendrer des réarrangements structuraux des autres domaines, la rendant inactive pour l'étape d'intégration. Une partie de mon travail de thèse s'est inscrite dans ce contexte et a ainsi consisté d'une part, à vérifier l'existence de cet assemblage sur le domaine isolé et sur la protéine entière en solution et d'autre part, à tester la capacité de molécules d'intérêt à se lier et à stabiliser ce « nouvel » assemblage, dans le but de tester leurs effets sur le cycle réplcatif rétroviral. Dans ce cadre, j'ai dû mettre en place au laboratoire différents protocoles et techniques complémentaires, aussi bien biochimiques que biophysiques, pour tenter de répondre à nos questions. Une autre partie de mon travail de thèse a consisté en l'étude structurale de l'intégrase entière du RAV-1, pour laquelle des données structurales n'étaient disponibles que pour des fragments de la protéine. Elle a nécessité la mise en place de protocoles de surproduction en système procaryote (*E. coli*) et de purification adaptés qui ont permis d'initier une caractérisation de la protéine en solution (état oligomérique, dispersité de l'échantillon), des essais de cristallisation et de réaliser une étude structurale à basse résolution par diffusion des rayons X aux petits angles (SAXS).

Une introduction bibliographique décrit tout d'abord le contexte scientifique de cette étude s'inscrivant dans le domaine de la rétrovirologie. Puis la méthodologie employée, ainsi que les résultats obtenus, sont présentés.

Comme ce premier projet avait demandé un fort travail de mise au point, mais n'avait pas abouti à de nouvelles données cristallographiques, j'ai été impliqué dans un deuxième projet portant sur l'étude structurale de deux α -galactosidases thermostables nommées AgaA et AgaB, produites par le microorganisme thermophile *Geobacillus stearothermophilus*. Cette étude est ainsi décrite dans la deuxième partie de ce manuscrit.

Débutée avant mon arrivée au laboratoire, cette étude a été initialisée en 2003 en collaboration avec l'équipe du Pr. Ralf Mattes (Institut de Génétique Industrielle, Université de Stuttgart, Allemagne) qui a assuré la préparation et la caractérisation des échantillons. Elle s'inscrit dans l'une des principales thématiques du laboratoire qui concerne l'étude structure/fonction de glycoside hydrolases. L'objectif du projet est de mieux comprendre les déterminants moléculaires et structuraux responsables des propriétés catalytiques différentielles des enzymes AgaA et AgaB, qui présentent pourtant une identité de séquence de 97 %. Lors de mon arrivée sur le projet, des données de diffraction des rayons X avaient déjà été obtenues pour AgaB (1,8 Å) et un mutant d'intérêt, le mutant AgaA A355E (2,8 Å), qui présente des propriétés catalytiques similaires à AgaB. Toutefois, le problème des phases n'avaient alors pas pu être résolu empêchant de remonter à l'information de structure. En 2010, la structure d'un homologue de ces enzymes a été résolue nous permettant de résoudre le problème des phases par remplacement moléculaire. Je me suis donc dans un premier temps consacré à l'affinement et à l'analyse de ces structures. Dans un second temps, il m'a fallu déterminer les conditions de cristallisation d'une forme tronquée de l'enzyme AgaA (ou AgaA-Δ9) avant de pouvoir résoudre sa structure cristalline à 3,2 Å de résolution. Au cours de ce projet, j'ai également pu résoudre la structure cristalline de l'enzyme AgaA A355E en complexe avec un inhibiteur (le 1-désoxygalactonojirimycine) ainsi que de deux mutants inactifs, AgaA A355E D548N et AgaA A355E D478A, respectivement en complexe avec les substrats D-raffinose (2,8 Å) et D-stachyose (3,6 Å).

Après une description générale des glycoside hydrolases et des α -galactosidases, la méthodologie employée ainsi que les résultats obtenus sont résumés et présentés sous la forme d'un article.

Publication sur le projet intégrase

Ballandras A, Moreau K, Robert X, Confort MP, **Merceron R**, Haser R., Ronfort C., Gouet P. (2011) A crystal structure of the catalytic core domain of an avian sarcoma and leukemia virus integrase suggests an alternate dimeric assembly. *PLoS One* 6: e23032.

Publication sur le projet α -galactosidase

Merceron R., Foucault M., Haser R., Mattes R., Watzlawick H., Gouet P. (2012) The molecular mechanism of thermostable alpha-galactosidases AgaA and AgaB explained by x-ray crystallography and mutational studies. *J Biol Chem* 287: 39642-39652.

PARTIE 1 :
ÉTUDE MOLÉCULAIRE ET STRUCTURALE D'UNE
INTÉGRASE RÉTROVIRALE POUR LE DÉVELOPPEMENT DE
NOUVEAUX ANTIRÉTROVIRAUX

EXPOSÉ BIBLIOGRAPHIQUE

1. Historique

Le terme « rétrovirus » décrit les virus de la famille des *Retroviridae*, qui comportent un génome diploïde à ARN simple brin (polarité positive) ; ces virus ont une forme sphérique avec un diamètre variant autour de 120 nm. Ces virus enveloppés utilisent au cours de leur cycle de réplication, une enzyme virale de type ADN polymérase ARN dépendante, la transcriptase inverse, qui catalyse la synthèse d'ADN double brin à partir du génome viral à ARN simple brin. Cet ADN rétroviral va pouvoir être intégré de façon stable dans le matériel génétique de la cellule infectée par une autre protéine virale, l'intégrase.

Les premières observations de ces virus remontent au début du vingtième siècle avec les travaux de W. Ellermann et O. Bang en 1908, portant sur des leucémies chez les poulets (Ellermann & Bang 1908). Les auteurs de ces travaux ont mis en évidence que ces pathologies pouvaient être transmises à des poulets sains à partir d'ultra-filtrats de cellules leucémiques provenant de poulets malades, démontrant ainsi l'origine virale de ces néoplasies dues à un virus appelé « *Avian Leukosis Virus* » ou ALV. Trois ans plus tard, la même observation fut faite par P. Rous, qui démontra la possibilité de transmettre des sarcomes chez des poulets sains à partir d'ultra-filtrats de cellules tumorales, et mis ainsi en évidence l'existence de l'agent infectieux *Rous Sarcoma Virus* ou RSV (Rous 1911). Ces virus infectant les espèces aviaires, et aujourd'hui regroupés sous le terme « *Avian Sarcoma and Leukemia Virus* » ou ASLV, ont été les premiers rétrovirus observés.

Plus tard, ces travaux sur les rétrovirus dits « oncogènes » furent étendus aux mammifères, avec la découverte d'agents rétroviraux responsables du carcinome mammaire et de leucémies chez la souris (Bittner 1936; Gross 1957). De nombreux autres rétrovirus induisant des leucémies et des sarcomes furent par la suite identifiés chez les poissons, les amphibiens, les insectes et les primates (Coffin *et al.* 1997). Ces virus ont été en conséquence décrits par le terme de « virus tumoraux à ARN » ou « oncornavirus ».

Le genre *Lentivirus* des *Retroviridae* regroupe des rétrovirus exogènes de mammifères (bovins, équidés, félins, caprins et primates) caractérisés par une longue période d'incubation avant l'apparition de pathologies. La première description d'une maladie à lentivirus fut faite au début du vingtième siècle avec les travaux de H. Vallée et H. Carrée en 1904 (Vallée & Carré 1904) portant sur l'anémie infectieuse du cheval, même si l'origine rétrovirale de cette pathologie ne fut comprise que plus tard. En 1954, les travaux de B. Sigurdsson mirent en évidence l'origine rétrovirale de la maladie de Visna touchant les moutons, et décrite comme une maladie à évolution lente (Sigurdsson 1954). En 1957, B. Sigurdsson isola et caractérisa ce rétrovirus qui fut le premier lentivirus véritablement décrit (Sigurdsson *et al.* 1957). La découverte de lentivirus infectant les singes et les chats vint par la suite (Letvin *et al.* 1985; Pedersen *et al.* 1987).

Une avancée majeure dans le domaine de la rétrovirologie fut la découverte en 1970 par deux équipes indépendantes, celle de H. Temin et S. Mizutani d'une part et celle de D. Baltimore d'autre part, de l'existence au sein de particules virales d'une enzyme possédant une activité de type ADN polymérase ARN dépendante appelée « transcriptase inverse » (Baltimore 1970; Temin & Mizutani 1970). Cette découverte, bien que n'étant pas une spécificité unique aux rétrovirus (cas des virus du groupe VII dans la classification de D. Baltimore, dont les *Hepadnaviridae*), permit de faciliter la reconnaissance de nouveaux rétrovirus et de renforcer l'hypothèse selon laquelle leur information génétique est portée par des molécules d'ARN simples brins au sein des virions et par des molécules d'ADN doubles brins au sein des cellules infectées. Elle permit également de remettre en question le dogme central de la biologie moléculaire énoncé par F. Crick en 1958 (Crick 1958) et en conséquence

actualisé en 1970 (Crick 1970) selon lequel l'information génétique ne pouvait être transmise que de l'ADN vers l'ARN. Cette découverte permit la création du terme de « rétrovirus » qui remplaça les termes jusqu'alors employés de virus « tumoraux à ARN » ou « oncogènes à ARN ».

Malgré l'ensemble de ces travaux sur des systèmes animaux, il n'était alors pas envisagé que ces rétrovirus puissent exister chez l'Homme et induire des pathologies similaires. L'existence possible de rétrovirus infectant l'espèce humaine fut toutefois rapportée dès 1971 avec l'isolement d'un virus, qui ne sera classifié qu'en 1987 comme étant un rétrovirus du genre *Spumavirus*, à partir d'un carcinome du nasopharynx chez l'Homme, sans pour autant établir son rôle étiologique direct (Achong *et al.* 1971). Plus tard fut découvert par l'équipe de R. Gallo un autre cas de rétrovirus infectant l'Homme, le rétrovirus T-lymphotropique humain de type 1 ou HTLV-1 (pour *Human T-Lymphotropic Virus type 1*) comme étant l'agent étiologique de leucémies à cellules T (Poiesz *et al.* 1980). Le HTLV-2 qui est moins pathogène fut découvert deux ans plus tard par la même équipe (Kalyanaraman *et al.* 1982).

L'apparition en 1981 d'une épidémie caractérisée par un Syndrome d'ImmunoDéficiency Acquis ou SIDA, et l'apparition de maladies opportunistes et de cancers chez certains individus appartenant à des communautés homosexuelles aux Etats-Unis, mena à la découverte d'un nouveau rétrovirus infectant l'espèce humaine. En 1983, un rétrovirus isolé à partir de patients atteints de lymphoadénopathies, fut découvert par F. Barre-Sinoussi, J. C. Chermann et L. Montagnier et nommé « virus associé à une lymphoadénopathie » ou LAV (pour *Lymphadenopathy Associated Virus*) (Barre-Sinoussi *et al.* 1983). Ce virus était capable de se répliquer et de détruire les lymphocytes T CD4+ en culture, permettant de supposer qu'il s'agissait de l'agent étiologique du SIDA. Un an plus tard, l'équipe de R. Gallo isola à partir du sang de patients atteints du SIDA, un rétrovirus qu'il nomma « rétrovirus T-lymphotropique humain de type 3 » ou HTLV-3 (pour *Human T-Lymphotropic Virus type III*) (Gallo *et al.* 1984). En 1985, il fut montré que les virus LAV et HTLV-3 étaient identiques (Marx 1985). L'année suivante, la communauté scientifique adopta le nom de « Virus de l'Immunodéficience Humaine de type 1 » ou VIH-1 en remplacement de LAV et HTLV-3 (Coffin *et al.* 1986). En 1986, un nouveau type de LAV appelé « LAV-2 » et renommé la même année « VIH-2 », fut découvert par l'équipe de L. Montagnier (Clavel *et al.* 1986).

Les virus VIH-1 et 2 sont aujourd'hui considérés comme les premiers et seuls rétrovirus du genre *Lentivirus* découverts dans l'espèce humaine. Le VIH de type 1 est désormais reconnu comme un problème majeur de santé mondiale avec plus de 34 millions de personnes infectées dans le monde et 1,7 millions de décès à la fin de l'année 2011 (www.unaids.org).

2. Taxonomie des rétrovirus

Les *Retroviridae* constituent une très large famille de virus. Ils appartiennent au groupe VI (virus à ARN avec intermédiaire à ADN) dans la classification générale des virus décrite par D. Baltimore (Baltimore 1971). La découverte de nouveaux rétrovirus a rendu nécessaire la mise en place de systèmes de classification basés sur différents critères.

Les virus de la famille des *Retroviridae* ont initialement été divisés en trois sous-familles en fonction de leur pathogénicité (Hong *et al.* 2010). La première sous-famille comprenait les *Oncovirinae* qui sont responsables de leucémies et d'affections malignes chez les individus infectés, parmi lesquels les virus du groupe ASLV infectant les espèces aviaires. La classification des *Oncovirinae* a par la suite été complétée en tenant compte des aspects morphologiques des particules virales observées au microscope électronique et a permis de définir les groupes A à D (Bernhard 1960). La deuxième sous-famille comprenait les *Lentivirinae* ou virus « lents »

responsables de pathologies à évolution lente caractérisées par une longue période d'incubation et par une infection persistante (exemple du VIH-1 responsable du SIDA chez l'Homme). La dernière sous-famille comprenait les *Spumaretrovirinae* considérés comme non pathogènes et qui provoquent *in vitro* l'apparition de cellules géantes multinucléées (*syncytia*) présentant de nombreuses vacuoles conduisant à une lyse cellulaire complète en quelques jours.

La taxonomie des rétrovirus a par la suite été redéfinie par le comité international de taxonomie des virus ou ICTV (pour *International Committee on Taxonomy of Viruses*) (King *et al.* 2011). La classification des rétrovirus par l'ICTV est principalement basée sur les similarités de séquences au sein du gène viral *pol*. Elle classe ces virus en deux sous-familles, les *Orthoretrovirinae* et les *Spumaretrovirinae*. Chaque sous-famille comprend différents genres avec les *Alpharétrovirus*, *Bétarétrovirus*, *Gammaretrovirus*, *Deltarétrovirus*, *Epsilon-rétrovirus* et *Lentivirus* au sein des *Orthoretrovirinae* et les *Spumavirus* au sein des *Spumaretrovirinae* (figure 1).

Les rétrovirus peuvent également être classifiés en deux grandes catégories qui sont les rétrovirus endogènes et les rétrovirus exogènes. Les rétrovirus exogènes se transmettent de façon horizontale d'un individu à l'autre et de façon verticale (de la mère à l'enfant, par exemple lors de l'accouchement ou de l'allaitement). La plupart de ces virus constituent des pathogènes pour différentes espèces. Contrairement aux rétrovirus exogènes, les rétrovirus endogènes sont intégrés dans le génome des cellules de la lignée germinale et peuvent être transmis de génération en génération.

Les rétrovirus endogènes ont été mis en évidence pour la première fois dans des cellules embryonnaires chez des poulets saints (Weiss 2006). On présume que ces derniers dérivent d'événements intégratifs de rétrovirus exogènes dans la lignée germinale d'un hôte spécifique. Ils ont été retrouvés chez l'Homme ainsi que chez de nombreux autres eucaryotes. Chez l'Homme, ils constituent près de 8 % du patrimoine génétique (Bannert & Kurth 2004). La plupart d'entre eux sont défectifs pour la réplication du fait de nombreuses délétions ou de mutations ponctuelles et ne sont pas transcriptionnellement actifs. Ils sont donc en majorité non infectieux, même si leur participation dans certains processus pathologiques comme certains cancers n'est pas écartée (Ruprecht *et al.* 2008; Brodziak *et al.* 2012). Ces rétrovirus pourraient toutefois participer activement à certains processus physiologiques, l'exemple le plus connu étant celui de la protéine d'enveloppe nommée « syncytine » exprimée par la famille des rétrovirus endogènes humains W où HERV-W (pour *Human Endogenous RetroVirus type W*). Cette protéine d'origine virale joue un rôle essentiel lors de la formation du placenta et la dérégulation de son expression serait associée à des cas de pré-éclampsie (Sugimoto & Schust 2009).

Enfin, les rétrovirus peuvent également être classifiés en fonction de la complexité de leur génome définissant ainsi les rétrovirus qualifiés de « simples » (comme les ASLV) qui contiennent dans leur génome uniquement des gènes codant des protéines dites « structurales » et « enzymatiques » et les rétrovirus qualifiés de « complexes » (comme le VIH-1) qui contiennent en plus de ces gènes, des gènes codant des protéines « régulatrices » et des protéines « accessoires ».

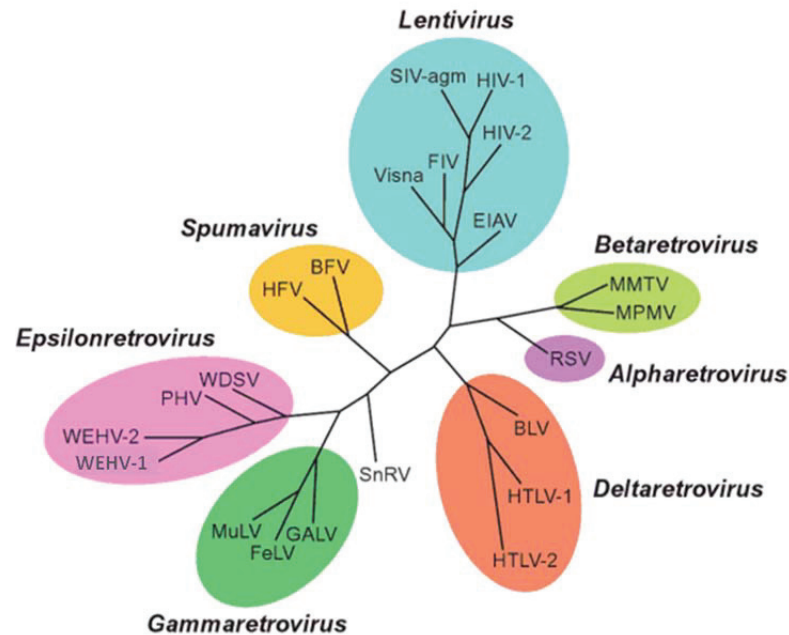


Figure 1 : Dendrogramme de classification des rétrovirus basé sur la conservation du gène viral *pol* d'après (King *et al.* 2011).

3. Organisation structurale et génomique

Malgré la grande variabilité des rétrovirus en matière de spécificité d'hôtes et d'interactions avec son hôte, ces derniers partagent des caractéristiques communes concernant la structure du virion et l'organisation du génome.

3.1. La particule virale

Les particules virales matures ou virions sont des particules sphériques d'une centaine de nanomètres de diamètre (**figure 2**). Elles possèdent une enveloppe lipidique provenant de la membrane plasmique des cellules infectées, formée suite au bourgeonnement de la particule. Cette enveloppe contient le plus souvent deux types de glycoprotéines : des glycoprotéines de surface (SU) et des glycoprotéines d'enveloppe transmembranaire (TM), ces dernières traversant la membrane, ainsi que des protéines membranaires issues de la cellule infectée.

Sous l'enveloppe, se trouvent les protéines de matrice (MA) assemblées sous forme sphérique, qui enferment une capside virale issue de l'assemblage des protéines de capside (CA). A l'intérieur, se trouvent la protéase virale (PR) et deux copies du génome viral à ARN simple brin associées à l'enzyme transcriptase inverse (RT). Associées au génome, se trouvent également la nucléocapside constituée de l'assemblage de nucléoprotéines (NC) et la protéine intégrase (IN) (King *et al.* 2011).

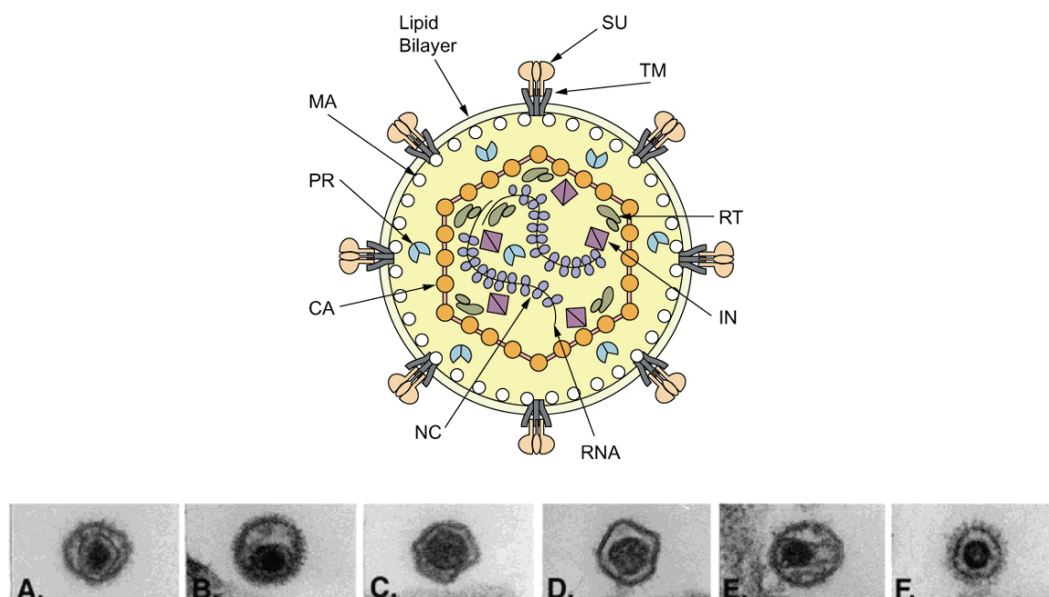


Figure 2 : (Haut) Vue schématique de la coupe transversale d'une particule rétrovirale d'après (Voisset & Andrawiss 2000).
(Bas) Morphologie des particules virales matures observée au microscope électronique pour les alpharétrovirus (A), bétarétrovirus (B), gammarétrovirus (C), deltarétrovirus (D), lentivirus (E) et spumavirus (F) d'après (King *et al.* 2011).

3.2. Le génome viral

3.2.1. Généralités

L'organisation et la structure du génome sont globalement similaires pour tous les rétrovirus. Le génome viral est constitué de deux molécules d'ARN monocaténaire à polarité positive de 7 à 13 kb et qui contiennent des modifications post-transcriptionnelles telles une coiffe à l'extrémité 5', une queue poly-adénylée à l'extrémité 3' et des méthylations sur certaines adénosines. Les deux molécules d'ARN sont assemblées de façon non covalente pour former un génome diploïde. En plus du génome viral à ARN, sont également retrouvées dans les virions matures d'autres molécules d'ARN comme des ARNt et des ARN ribosomiques cellulaires. Les ARNt (pour ARN de transfert) retrouvés ont un rôle crucial dans l'initiation de la transcription inverse car ils servent d'amorce pour la synthèse de l'ADN. Des molécules d'ADN sont également retrouvées dans certaines particules virales en faible quantité et seraient le produit de l'initiation précoce de la transcription inverse dans les virions. Chez les virus du genre *Spumavirus*, près de 20 % des particules libérées contiennent déjà un génome sous forme d'ADN double brin du fait de l'activité de transcription inverse qui constitue une étape tardive du cycle réplcatif rétroviral chez ces virus (King *et al.* 2011).

3.2.2. Structure

Selon la complexité de leur génome, on distingue deux groupes de rétrovirus : les rétrovirus « simples » et les rétrovirus « complexes ».

Pour les rétrovirus à génome simple (cas des alpha, bêta, epsilon et gammarétrovirus), leur génome est composé de trois gènes qui sont :

- le gène *gag* (pour *group-specific antigen*), qui code les protéines « structurales » de la particule virale (CA, MA, NC),

- le gène *pol* (pour *polymerase*), qui code les protéines « enzymatiques » impliquées dans la réplication et l'intégration virale (IN, RT, PR), la protéase PR pouvant selon les virus être codée par le gène *gag*,
- le gène *env* (pour *envelope*), qui code les glycoprotéines de l'enveloppe virale (TM, SU) (**figure 3**).

Pour les rétrovirus à génome complexe (cas des deltavirérovirus, spumavirus et lentivirus) : en plus des trois gènes *gag*, *pol*, *env*, leurs génomes codent des protéines additionnelles (protéines « régulatrices » et protéines « accessoires »).

A l'état d'ADN complémentaire ou de provirus (forme intégrée), le génome viral est encadré par deux régions non codantes, appelées « LTR » (pour *Long Terminal Repeat*) impliquées dans l'intégration et la régulation de l'expression du génome viral. Ces séquences se divisent en trois éléments : une région U3 suivie de la région R (pour *Repeat*) puis de la région U5 (**figure 3**).

La région U3 contient à son extrémité 5' le site d'attachement de l'intégrase nommé « *att* » (pour *attachment*). Ce site est reconnu spécifiquement par la protéine intégrase. Cette région contient également des signaux qui sont reconnus par la machinerie transcriptionnelle cellulaire. Une séquence nommée « *Enhancer* », dans laquelle se situent des sites de fixations des facteurs de transcription, permet de réguler les niveaux d'expression de l'ARN viral. On trouve également dans la région U3 la boîte CAAT, élément régulateur de la boîte TATA localisée 25 nucléotides en amont du site d'initiation de la transcription (jonction des régions U3 et R) et à laquelle se lie le facteur de transcription TFIID, dirigeant la fixation de l'ARN polymérase cellulaire de type II.

La région R intervient lors de la transcription inverse. Cette région contient également le site nécessaire pour l'ajout de la coiffe (site « *CAP* ») sur l'ARN viral et le signal de polyadénylation (site « *PolyA* »).

La région U5 contient, comme la région U3, une partie de la séquence *att* reconnue par l'intégrase (**figure 3**).

3.2.3. L'ARN génomique et les ARN messagers viraux

Pour les rétrovirus « simples » (comme les virus de type ALV), l'expression du provirus conduit à la production de deux ARN messagers : un transcrit primaire appelé « ARN génomique » et un transcrit mono-épissé appelé « ARN sous-génomique » ou « ARNm *env* ». Pour les rétrovirus complexes, il y a en plus, des transcrits multi-épissés permettant la traduction des protéines « accessoires » et des protéines « régulatrices ». L'ARN génomique débute par les régions R-U5 et se termine par les régions U3-R.

Sur les ARNm (pour ARN messenger) viraux, se trouvent de nombreuses séquences cis-régulatrices nécessaires à l'expression du génome viral (séquences d'épissage, séquences IRES,...). Nous pouvons citer en particulier les suivantes : en aval de la région U5 se trouve une séquence PBS (pour *Primer Binding Site*) qui sert de site de fixation pour un ARN de transfert à partir duquel s'amorce la transcription inverse. On observe également une région Ψ correspondant au signal d'encapsidation de l'ARN viral génomique dans les particules en formation. En amont de la région U3, se trouve une séquence riche en A/G appelée « PPT » (pour *PolyPurine Tract*) qui, lors de la transcription inverse, résiste à l'action de la RNase H et sert d'amorce ARN pour la synthèse du second brin d'ADN (**figure 3**). Chez les lentivirus, cette séquence PPT est aussi présente en position centrale dans le génome.

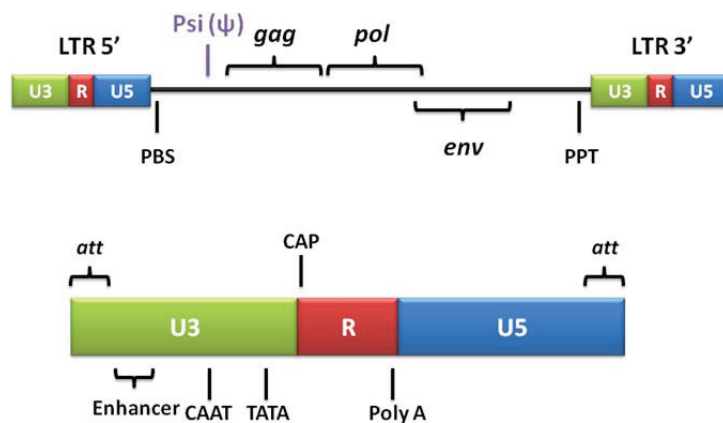


Figure 3 : (Haut) Organisation génomique générale des provirus.
(Bas) Structure générale des LTR rétroviraux.

3.3. Les protéines virales enzymatiques et structurales

A partir des deux ARN messagers, trois précurseurs polyprotéiques (Pr) sont produits : Pr Gag, Pr Gag-Pol et Pr Env.

L'ARNm génomique permet la traduction de deux polyprotéines : le Pr Gag et le Pr Gag-Pol. En 3' du gène *gag* se trouve un codon stop qui est, dans 90 % des cas, interprété correctement par le ribosome et conduit à la synthèse du précurseur polypeptidique Pr Gag. L'environnement du codon stop permet, dans 10 % des cas, un décalage du cadre de lecture et la synthèse d'un précurseur polypeptidique Pr Gag-Pol.

Les précurseurs polypeptidiques Pr Gag et Pr Gag-Pol, une fois clivés par la protéase virale (PR), donnent trois protéines majeures qui sont la capside (CA), la matrice (MA) et enfin, la nucléocapside (NC).

Le précurseur polypeptidique Pr Gag-Pol permet l'obtention des différentes enzymes virales après clivage du Pr Gag-Pol par la protéase virale. Ces enzymes sont : la transcriptase inverse (RT) et l'intégrase (IN). La protéase (PR) est, quant à elle, codée soit par le gène *gag* soit par le gène *pol* selon le type de rétrovirus (*gag* pour les ALV et *pol* pour le VIH-1 par exemple).

L'ARNm *env* permet la traduction d'un précurseur polypeptidique Pr Env qui, après glycosylation, est clivé par une protéase cellulaire pour générer les glycoprotéines de l'enveloppe virale : la TM et la SU.

3.3.1. Cas des rétrovirus complexes

Les rétrovirus complexes possèdent, en plus des gènes *gag*, *pol* et *env*, des gènes supplémentaires codant des protéines « régulatrices » et des protéines « accessoires ».

Les protéines « régulatrices » jouent un rôle important dans l'interaction hôte-pathogène et vont avoir un impact significatif sur la réplication virale (Karn & Stoltzfus 2012). Dans le cas du VIH-1, le génome de ce rétrovirus comprend deux gènes codant des protéines « régulatrices », Tat et Rev. La protéine Tat (pour *TransActivator of transcription*) est un transactivateur puissant de la

transcription du VIH-1. La protéine Rev (pour *Regulator of Expression of the virion*) favorise l'export nucléaire des ARN messagers viraux non et mono-épissés vers le cytoplasme.

Le génome du VIH-1 contient aussi des séquences codant des protéines « accessoires » telles que Nef, Vif, Vpr ou Vpu (Li *et al.* 2005; Malim & Emerman 2008; Gramberg *et al.* 2009) (**figure 4**). Ces protéines « accessoires » sont nécessaires dans certains types cellulaires ou dans certaines conditions d'infection :

- La protéine Nef (pour *Negative regulator factor*) possède quatre activités indépendantes au cours de la réplication. Elle affecte l'expression de certaines protéines cellulaires de surface, interfère avec des signaux de transduction cellulaire et augmente l'infectiosité du virion ainsi que la réplication.
- La protéine Vif (pour *Virus infectivity factor*) est essentielle pour la réplication du virus dans les lymphocytes du sang périphérique, les macrophages et certaines lignées comme les cellules non-permissives. Vif se fixe au facteur cellulaire APOBEC3G (pour *APOLipoprotein B mRNA editing Enzyme, Catalytic polypeptide-like 3G*). L'enzyme cellulaire APOBEC3G constitue l'un des mécanismes de défense antivirale naturelle des cellules de mammifères contre les rétrovirus. APOBEC3G modifie la composition génétique du virus en désaminant l'ARN viral ce qui forme des virus non infectieux. Le rôle essentiel de la protéine Vif est de neutraliser cette ligne de défense en provoquant la dégradation d'APOBEC3G, permettant ainsi la propagation du virus.
- La protéine Vpr (pour *Viral protein R*) possède de nombreuses fonctions pour la réplication du virus notamment dans l'import nucléaire du complexe de pré-intégration et de l'apoptose, dans l'arrêt du cycle cellulaire en phase G2 et dans la transactivation des LTR viraux et de certains gènes cellulaires.
- La protéine Vpu (pour *Viral protein U*) augmente la libération des nouvelles particules virales et induit la dégradation des récepteurs CD4. Elle est capable de bloquer l'action de la Téthérine, une protéine cellulaire qui retient le virus à la surface de la cellule.

3.3.2. Cas des oncovirus

Chez certains virus, en particulier chez les oncovirus, d'autres gènes sont présents dans le génome viral et sont souvent très similaires à certains gènes cellulaires et en particuliers à des proto-oncogènes. Dans ces oncovirus, le matériel génétique supplémentaire va souvent pouvoir induire une modification de gènes de « structure » par insertion voire même par leur remplacement. Ces virus peuvent donc présenter des délétions dans les gènes *gag*, *pol* ou *env* et devenir défectifs pour la réplication (cas de certaines souches du RSV dont la souche *Bryan high-titer*) (**figure 4**). Ces derniers nécessitent alors la présence d'un virus dit « complément » apportant en *trans* les fonctions nécessaires à leur réplication (exemple du *Rous Associated Virus* ou RAV, virus complément du RSV).

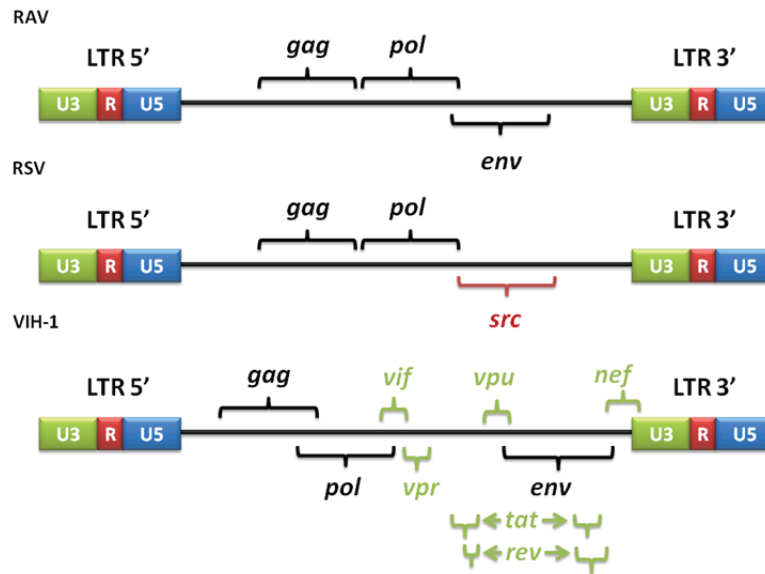


Figure 4 : Organisation génomique des provirus de trois rétrovirus représentatifs. **(Haut)** Provirus d'un rétrovirus « simple » non defectif (exemple : ALV, RAV). **(Milieu)** Provirus d'un rétrovirus « simple » defectif contenant l'oncogène SRC (exemple de l'oncovirus RSV souche *Bryan high-titer*). **(Bas)** Provirus d'un rétrovirus « complexe » (lentivirus VIH-1).

4. Assemblage et réplication des particules virales

Le cycle de réplication des rétrovirus est divisé en deux grandes phases (**figure 5**). La première phase (phase pré-intégrative ou « précoce ») est initiée par la reconnaissance de la cellule cible par le virion mature et comprend tous les processus aboutissant à l'intégration de l'ADN viral dans l'ADN génomique. La deuxième phase (phase post-intégrative ou « tardive ») commence avec l'expression du provirus et comprend tous les processus allant jusqu'au bourgeonnement et à la maturation des virions (King *et al.* 2011).

4.1. Phase pré-intégrative

Comme pour la plupart des virus, l'infection par les rétrovirus nécessite une interaction spécifique de la protéine d'enveloppe virale SU avec un récepteur de la surface cellulaire. Le récepteur cellulaire pour le VIH-1 est une molécule de la superfamille des immunoglobulines appelée « cluster de différenciation 4 » (CD4) permettant au virus d'infecter les lymphocytes T CD4⁺ et les macrophages CD4⁺ (Dalglish *et al.* 1984). L'interaction de l'enveloppe virale avec son récepteur spécifique n'est pas toujours suffisante pour initier la fusion membranaire. Chez le VIH-1, la protéine d'enveloppe GP120 doit interagir avec son récepteur spécifique CD4 puis avec un corécepteur aux chimiokines (CCR5 ou CXCR4) qui sont des protéines de la superfamille des récepteurs couplés aux protéines G (Feng *et al.* 1996).

La liaison du virus au récepteur et co-récepteur cellulaire induit un changement de conformation des glycoprotéines d'enveloppe, permettant d'initier la fusion du virus avec la membrane cellulaire et son entrée dans la cellule hôte (Blumenthal *et al.* 2012).

La transcription inverse est catalysée dans le cytosol (à l'exception des spumavirus où la réaction est initiée dès l'assemblage et le bourgeonnement des virions) par la transcriptase inverse virale contenue dans un complexe multi-protéique contenant différentes protéines virales et

cellulaires, selon un mécanisme semblable chez tous les rétrovirus (Warrilow *et al.* 2009). Au cours de cette réaction, le génome viral à ARN simple brin va être converti en ADN double brin. Le manque de fiabilité de la transcriptase inverse est une des raisons de la grande variabilité génétique observée chez certains rétrovirus et en particulier chez le VIH-1. C'est en partie pour cela que ce dernier est capable d'échapper rapidement au système immunitaire de l'hôte et de développer des résistances aux drogues antivirales (Le Grice 2012).

Une fois synthétisé, l'ADN viral est transporté vers le noyau au sein du complexe de pré-intégration ou PIC (pour *Pre-Integration Complex*) contenant différentes protéines aussi bien virales que cellulaires (Krishnan & Engelman 2012). La cassure de la membrane nucléaire lors de la mitose semble parfois nécessaire à l'entrée du PIC dans le noyau (nécessité d'infecter des cellules en division) tandis que certains rétrovirus (comme le VIH-1 ou le RSV) sont capables d'infecter des cellules en division ainsi que des cellules quiescentes (Yamashita & Emerman 2006).

Après l'entrée du PIC dans le noyau, la protéine virale intégrase catalyse l'insertion de l'ADN viral linéaire bicaténaire dans le génome de la cellule hôte. L'ADN viral intégré est appelé « provirus » (Krishnan & Engelman 2012).

4.2. Phase post-intégrative

Après son intégration dans le génome de la cellule hôte, le provirus va servir de matrice pour la synthèse des ARN viraux, qui vont coder toutes les protéines « structurales » et « enzymatiques » et éventuellement pour les rétrovirus « complexes », les protéines « régulatrices » et les protéines « accessoires ». La transcription, réalisée par l'ARN polymérase II cellulaire est initiée au niveau de la jonction U3-R du LTR 5' et se termine à la jonction R-U5 du LTR 3'.

Suite à la transcription du provirus, une partie des ARNm viraux sera directement dirigée dans le cytoplasme et empaquetée dans les virions. Suite à la synthèse des protéines virales, le processus de formation de la particule virale ou assemblage est initié. Pour la plupart des virus, l'assemblage des capsides et le bourgeonnement se font simultanément. L'assemblage semble impliquer l'interaction des précurseurs Gag et Gag-Pol avec le génome ainsi qu'avec d'autres précurseurs Gag et avec la membrane cellulaire. On observe ensuite le bourgeonnement de ces virions immatures, puis la maturation au cours de laquelle la protéase virale réalise le clivage des précurseurs Gag et Gag-Pol (Lee *et al.* 2012). Chez le VIH-1, le résultat le plus visible dans la maturation est l'apparition d'une structure conique correspondant à la protéine de capside CA autour du complexe ARN-protéines. Les virions obtenus sont dits matures et sont alors infectieux.

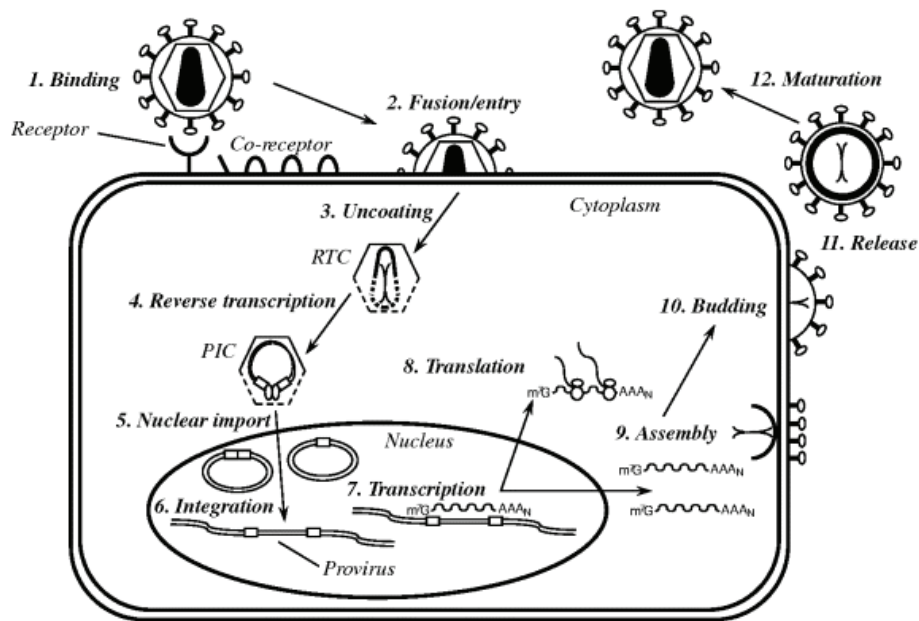


Figure 5 : Représentation schématique du cycle réplcatif rétroviral d'après (Hefferon 2006).

5. Pathogénèses rétrovirales

Bien qu'une partie des rétrovirus n'apparaisse pas reliée à l'apparition de pathologies après l'infection de l'hôte, ce qui semble être le cas pour une majorité des rétrovirus endogènes, les rétrovirus exogènes ont été décrits pour la première fois par leurs effets biologiques et en particulier pathogènes sur leurs hôtes.

Ces effets, qui varient fortement selon les rétrovirus et leurs hôtes, peuvent aller d'une simple virémie jusqu'à l'apparition plus ou moins rapide de processus tumoraux (cas des rétrovirus oncogènes comme ASLV), de syndromes d'immunodéficiences associés à diverses pathologies neurologiques comme la démence (cas des lentivirus comme VIH ou Visna) (Kurth & Bannert 2010).

5.1. Les rétrovirus dans l'oncogénèse

De nombreuses évidences ont permis de relier les rétrovirus aux processus malins (Maeda *et al.* 2008). Une majorité de ces rétrovirus a d'ailleurs été isolée pour la première fois à partir de tissus tumoraux chez des individus infectés. Parmi ces rétrovirus dits « oncogènes », on peut distinguer deux catégories. La première catégorie correspond aux oncovirus dits « rapides » qui déclenchent des tumeurs fulgurantes avec des effets visibles sur les cellules infectées. La deuxième catégorie correspond aux oncovirus dits « lents », qui déclenchent des pathologies malignes beaucoup plus lentement (plusieurs années), et qui n'induisent aucun effet visible sur les cellules infectées (Maréchal & Piolot 1999; Kurth & Bannert 2010).

La première catégorie regroupe des oncovirus souvent désignés par les termes d'oncovirus « transducteurs » ou de virus à « pouvoir oncogène rapide ». La plupart de ces virus apparaissent comme possédant dans leur génome des copies de parties de proto-oncogènes cellulaires. Ils dérivent leur potentiel oncogénique du fait de la surexpression de ces oncogènes viraux sous le contrôle du promoteur fort du LTR 5', induisant une perturbation de certaines voies de signalisation

cellulaires souvent impliquées dans le contrôle de la prolifération cellulaire. On peut citer pour exemple le cas de certaines souches du RSV qui possèdent à la place du gène viral *env* l'oncogène *src*. L'oncogène *src* code une protéine kinase qui va être surproduite dans les cellules infectées. La surproduction de cette protéine va induire une prolifération cellulaire incontrôlée et ainsi favoriser l'apparition de tumeurs (Stehelin *et al.* 1976; Brugge & Erikson 1977).

La deuxième catégorie regroupe des oncovirus souvent désignés par les termes d'oncovirus « cis-activateurs » ou d'oncovirus à « pouvoir oncogène lent ». Ces virus ne possèdent pas d'oncogènes dans leur génome et sont donc répliatifs. Un exemple concerne les virus de la leucose aviaire ou ALV qui peuvent induire avec une période de latence de près de six mois des lymphomes à cellules B chez les poulets. Il a été observé que les cellules tumorales issues des ALV présentaient un provirus inséré en un même endroit du génome et en particulier en amont d'un proto-oncogène cellulaire appelé « *c-myc* » (Hayward *et al.* 1981). L'insertion de ces provirus va induire le contrôle de l'expression du proto-oncogène *c-myc* par le LTR 3' du provirus, induisant une perturbation de la régulation de l'expression de ce proto-oncogène cellulaire dans le sens d'une surexpression. On parle d'insertion activationnelle ou de cis-activation. Cette surexpression va induire, comme dans le cas du RSV, la multiplication clonale des précurseurs des lymphocytes B chez les individus infectés. L'intégration du génome viral dans l'ADN cellulaire peut aussi inactiver l'expression d'un gène exerçant un contrôle négatif sur la progression du cycle cellulaire et induire la formation de tumeurs (cas du gène *p53* et de la souche *Friend* du virus MLV) (Prasher *et al.* 2001).

Un autre mécanisme utilisé par certains virus oncogènes est le mécanisme de trans-activation. Ce mécanisme utilisé entre autre par les rétrovirus HTLV-1 ou VIH-1 dits « trans-activateurs », implique certaines protéines virales comme Tax (pour *Transactivator of pX*) chez HTLV-1 et Tat chez VIH-1. Ces protéines vont pouvoir diffuser à travers les membranes cellulaires, augmenter l'expression du provirus mais également suractiver l'expression de certains gènes cellulaires contribuant ainsi à l'apparition de processus tumoraux (Grassmann *et al.* 2005; Nunnari *et al.* 2008).

5.2. Les rétrovirus dans l'immunodéficience

Certains rétrovirus, en particulier ceux appartenant au genre *Lentivirus*, sont associés à des syndromes d'immunodéficience chez les individus infectés, l'exemple le plus connu étant celui du lentivirus VIH-1, l'agent étiologique du SIDA chez L'Homme. La pathogénèse du VIH-1 est complexe et se caractérise principalement par une cytopathogénèse des lymphocytes T CD4⁺ matures et des macrophages aboutissant à l'apparition progressive d'une immunodéficience qui permettra par la suite l'installation de maladies dites « opportunistes ».

La déplétion des cellules CD4⁺, cellules orchestres du système immunitaire, est provoquée par différents mécanismes encore mal connus. Tout d'abord, la réplication du VIH peut induire à lui seul un effet cytopathogène sur les cellules infectées du fait de la toxicité directe des protéines virales surproduites ou encore des modifications des propriétés membranaires des cellules infectées par les protéines d'enveloppe et par le bourgeonnement des particules virales. Une autre origine à la cytopathie de ces cellules est la formation de *syncytia* de cellules CD4⁺ qui résultent de la fusion de plusieurs cellules CD4⁺ non infectées avec une cellule CD4⁺ infectée par le biais d'interactions CD4-gp120 (Swanstrom & Coffin 2012).

Le système immunitaire de l'hôte peut également contribuer à la destruction de ces cellules par les lymphocytes T CD8⁺ cytotoxiques, les cellules NK (pour *Natural Killer*) ou encore par le mécanisme de cytotoxicité cellulaire médiée par les anticorps. Il a également été montré que le VIH-1 était capable d'induire l'apoptose des cellules infectées (Swanstrom & Coffin 2012).

6. Restriction intracellulaire de l'infection

Les virus utilisent de nombreux facteurs cellulaires pour assurer un cycle réplcatif efficace. A l'inverse, les cellules de l'hôte expriment différentes protéines permettant de limiter ou supprimer la réplication virale. Ces protéines sont appelées « facteurs de restriction » et constituent un élément important de la réponse immunitaire innée contre l'infection virale (Malim & Bieniasz 2012).

Pour le VIH-1, ces protéines et leurs rôles dans les mécanismes de défense antiviraux sont très étudiés et quatre facteurs de restriction ont été identifiés à ce jour : TRIM 5 α (pour *Tripartite-Motif-Containing 5 α*), APOBEC3G (pour *APOLipoprotein B mRNA-editing Enzyme, Catalytic polypeptide-like 3G*), Tetherin (Téthérine, CD317) et SAMHD1 (pour *Sterile Alpha Motif and HD domain 1*) :

- TRIM 5 α interfère avec les étapes précoces du cycle réplcatif rétroviral et plus particulièrement avec l'étape de décapsidation. Des études ont montré qu'il était capable d'interagir avec la capsid virale perturbant l'étape de décapsidation et empêchant ainsi la transcription inverse et le transport du génome viral vers le noyau (Sastri & Campbell 2011).
- Les protéines de la famille APOBEC constituent une famille d'enzymes à activité cytidine désaminase qui convertissent les cytosines en uraciles au sein des ARN. Ces enzymes cellulaires introduisent donc des mutations lors de la rétrotranscription de l'ARN viral, aboutissant à la synthèse d'ADN viraux hypermutés. Ces mutations conduisent soit à la dégradation de ces ADN viraux, soit à l'intégration de provirus incapables de synthétiser des protéines fonctionnelles (Adolph *et al.* 2013).
- La Téthérine est une protéine associée aux membranes et qui agit sur la phase tardive du cycle réplcatif. Il a été montré que cette protéine permettait la rétention des particules virales bourgeonnantes à la surface des cellules les empêchant ainsi d'infecter les cellules voisines (Neil *et al.* 2008).
- SAMHD1 est une enzyme qui possède une activité désoxynucléoside triphosphate triphosphohydrolase qui va permettre la diminution des concentrations intracellulaires en nucléoside dans les cellules. Il a été montré que la diminution en nucléosides limitait la transcription inverse du génome viral et la réplication des virus dans les cellules myéloïdes (Lagette *et al.* 2011).

Afin d'assurer leur réplication, les virus ont toutefois élaboré des stratégies permettant de contrecarrer certains de ces facteurs de restriction. Pour le VIH-1, on peut citer pour exemple les protéines accessoires Vif et Vpu qui contrecarrent respectivement l'activité d'APOBEC3G et de la Téthérine. Il en est de même pour le VIH-2 avec la protéine accessoire Vpx qui contrecarre l'activité de SAMHD1 dans les cellules myéloïdes (Strebel *et al.* 2009).

7. SIDA et stratégies antirétrovirales

La prévention des infections rétrovirales et leur contrôle dans des populations infectées sont d'une importance majeure. Certaines infections rétrovirales s'accompagnent de pathologies sérieuses chez de nombreuses espèces, dans le domaine de l'agriculture (rétrovirus ASLV chez les volailles, BLV chez les bovins, EIAV chez les équidés), chez les animaux de compagnie (FeLV et FIV

chez les chats) ou encore chez l'Homme (HTLV-1 et VIH-1). Du fait de son caractère pandémique, le SIDA causé par le rétrovirus VIH-1 constitue un enjeu majeur de la recherche en rétrovirologie.

Des progrès scientifiques et thérapeutiques ont été réalisés afin de réduire le nombre total de nouveaux cas annuels d'infections et de décès liés au VIH-1. Cependant, le nombre total de personnes infectées continue de s'accroître, malgré une stabilisation de la prévalence de la maladie, en raison de l'allongement de la durée de vie des patients séropositifs (grâce aux progrès thérapeutiques) et du nombre annuel de nouvelles infections (lié à l'absence de traitement préventif comme un vaccin anti-VIH).

De nombreuses études portant sur les protéines virales, mais aussi sur son matériel génétique, ont été réalisées et ont permis la mise en place de stratégies antirétrovirales. Toutefois, la connaissance des mécanismes de réplication des rétrovirus souligne les difficultés majeures rencontrées dans la lutte contre les lentivirus et en particulier le VIH-1. Tout d'abord, l'intégration du génome viral dans le génome des cellules infectées permet aux rétrovirus de se comporter comme un élément génétique de l'hôte quand il est sous la forme de provirus. De cette façon, l'information génétique portée par le provirus va pouvoir être transmise aux cellules filles lors de la mitose. Les composés dirigés contre le VIH vont cibler essentiellement les virus réplicatifs permettant de réduire temporairement la charge virale. De plus, le VIH va pouvoir infecter les lymphocytes T CD4+ ainsi que les macrophages dans l'ensemble de l'organisme dont le système nerveux central. L'accès des composés antirétroviraux à l'ensemble des cellules infectées constitue une difficulté majeure et est pourtant nécessaire pour éviter l'existence de réservoirs à rétrovirus inaccessibles aux composés thérapeutiques. Enfin, la difficulté majeure réside dans la génération par la transcriptase inverse virale de multiples variants génétiques du fait des nombreuses erreurs engendrées par cette enzyme et du taux élevé de réplication virale, aboutissant très rapidement à l'émergence et au maintien de variants résistants aux composés antirétroviraux.

En dehors des recherches sur les thérapies géniques antirétrovirales (ARN antisens, ribozymes, ARN mimétiques...), la principale thérapie antirétrovirale actuelle est une thérapie chimique médicamenteuse constituée de composés qui vont inhiber la fonction des enzymes virales afin de perturber le cycle de réplication des rétrovirus (Arts & Hazuda 2012). De nombreux composés ont tout d'abord été découverts par des méthodes de criblage larges. Cependant, la découverte de nouveaux composés et la nécessité de leur haute spécificité, afin de réduire les effets secondaires liés au traitement et l'émergence de virus résistants, supposent un développement rationnel impliquant la connaissance de la structure tridimensionnelle des enzymes virales ciblées.

La monothérapie fut la première thérapie chimique antirétrovirale utilisée après la découverte en 1987 que l'AZT (agent antiviral utilisé par ailleurs) fonctionnait contre le VIH-1. Cette molécule, un analogue de la thymidine, permet d'inhiber la transcriptase inverse virale avec toutefois de nombreux effets secondaires et une efficacité jugée par la suite insuffisante. Dans les années 1995, apparut la bithérapie qui combine deux inhibiteurs de la transcriptase inverse et qui sera reconnue comme plus efficace. En 1996, l'arrivée d'une nouvelle classe de molécules, les antiprotéases virales, permit l'apparition de la trithérapie ou HAART (pour *Highly Active AntiRetroviral Therapy*). Celle-ci comprenait un inhibiteur de protéase virale ou un inhibiteur non nucléosidique de la transcriptase inverse en combinaison avec deux inhibiteurs nucléosidiques de la transcriptase inverse.

Ces dernières années, deux nouvelles classes de molécules sont apparues, les inhibiteurs d'entrée et de fusion ainsi que les inhibiteurs de l'intégrase. La combinaison de la trithérapie avec ces deux nouvelles classes de composés a donné naissance à la multithérapie.

Les effets secondaires associés à la multithérapie restent très nombreux et l'émergence de nouvelles mutations rendant le VIH-1 résistant aux drogues, rendent nécessaire le développement de nouveaux composés antirétroviraux (Cohen 2002). L'enzyme virale intégrase constitue une cible thérapeutique intéressante. Il a en effet été montré que l'intégrase rétrovirale possédait peu d'homologues cellulaires directs chez l'Homme (si ce ne sont les recombinases RAG) permettant de réduire les risques d'effets secondaires associés aux inhibiteurs de cette enzyme. Les études sur les inhibiteurs de l'intégrase ont commencé il y a près d'une dizaine d'années et actuellement, un seul inhibiteur de l'intégrase, le Raltégravir (médicament Isentress), a obtenu l'approbation de mise sur le marché en 2007 de la *Food and Drug Administration* ou FDA (**figure 6**). Des études biochimiques et structurales de cette enzyme sont nécessaires pour permettre de guider la conception rationnelle d'inhibiteurs spécifiques qui pourront être utilisées en complément de la multithérapie actuelle (Hazuda 2012; Malet *et al.* 2012).

Inhibiteurs de la transcriptase inverse	Inhibiteurs de la protéase	Inhibiteurs de la fusion et de l'entrée	Inhibiteurs de l'intégrase
Combivir	Agenerase	Fuzeon	Isentress
Emtriva	Aptivus	Selzentry	
Epivir	Crixivan		
Epzicom	Fortovase		
Hivid	Invirase		
Retrovir	Kaletra		
Trizivir	Lexiva		
Truvada	Norvir		
Videx EC	Prezista		
Videx	Reyataz		
Viread	Viracept		
Zerit			
Ziagen			
Edurant			
Intelence			
Rescriptor			
Sustiva			
Viramune			
Viramune XR			

Figure 6 : Les inhibiteurs anti-VIH-1 approuvés par l'administration américaine des denrées alimentaires et des médicaments (FDA) en 2013. Les inhibiteurs nucléosidiques et non nucléosidiques de la transcriptase inverse sont représentés respectivement dans les cases gris foncé et gris clair.

II. L'INTÉGRASE

1. Intégration et réplication

L'étape d'intégration, catalysée par l'enzyme virale intégrase, est une étape clé du cycle réplcatif rétroviral et commune à l'ensemble des rétrovirus. Seuls quelques auteurs ont mis en évidence une réplication rétrovirale en absence de processus d'intégration mais peu d'exemples similaires ont été rapportés jusqu'à présent (Panganiban & Temin 1983; Prakash *et al.* 1992).

L'intégration du génome viral dans le matériel génétique de la cellule hôte permet la transmission du provirus aux cellules filles lors de la division cellulaire. En effet, l'ADN viral sous forme épisomique, c'est-à-dire non intégré n'est pas capable de se répliquer de façon autonome. L'intégration du génome viral semble également nécessaire pour une expression efficace du provirus et à sa réplication. Il a en effet pu être montré que l'ADN viral non intégré ne permettait pas à lui seul d'assurer une réplication efficace des virions (Sloan & Wainberg 2011). L'intégration constitue donc une étape décisive dans le cycle de réplication des rétrovirus. Cependant, des formes non intégrées

des ADN viraux sont détectées dans l'ensemble des cellules infectées et pourraient participer à la réplication efficace des virions par différents mécanismes encore mal connus (rôle de réservoir) (Sloan & Wainberg 2011). Il est intéressant de noter que le processus d'intégration est également retrouvé chez d'autres virus comme ceux appartenant à la famille des *Parvoviridae* et des *Hepadnaviridae*.

2. Historique

L'enzyme virale intégrase a été détectée pour la première fois en 1978 chez le virus de la myéloblastose aviaire ou AMV (pour *Avian Myeloblastosis Virus*) et était alors définie comme une endonucléase non spécifique avec une masse moléculaire apparente de 32 kDa (enzyme p32) (Grandgenett *et al.* 1978). Il a par la suite été montré que cette enzyme clivait les séquences mimant les extrémités LTR du génome viral (Duyk *et al.* 1983). Un an plus tard, l'évidence fut faite que l'intégrase était nécessaire à la réplication rétrovirale et suffisante pour le processus d'intégration par l'analyse de certains mutants de l'intégrase du virus de la leucémie murine MuLV (pour *Murine Leukemia Virus*), qui étaient défectifs pour l'intégration et, conséquemment, la réplication (Schwartzberg *et al.* 1984). La même observation fut faite la même année chez le virus RSV (Hippenmeyer & Grandgenett 1984). La détermination du rôle de cette protéine dans le mécanisme d'intégration, aboutit à la mise en place d'une nomenclature standardisée pour l'appellation de cette enzyme, qui deviendra « l'intégrase rétrovirale » (Leis *et al.* 1988).

3. Propriétés fonctionnelles

3.1. Intégrase et transposases

L'intégrase catalyse l'intégration concertée de l'ADN viral bicaténaire rétro-transcrit au sein de l'ADN génomique de la cellule infectée. L'intégration rétrovirale est dite « concertée » car elle permet l'insertion des deux extrémités virales en un même site de l'ADN cellulaire. Les étapes de l'intégration sont catalysées par une triade catalytique appelée « motif D,D(35)E ». Ce motif comprend deux aspartates (D) et un glutamate (E), ce dernier étant séparé du deuxième aspartate par trente-cinq résidus. Ce motif est conservé au sein de toutes les intégrases rétrovirales (Jaskolski *et al.* 2009).

Le mécanisme par lequel les rétrovirus intègrent l'ADN viral dans l'ADN génomique est retrouvé chez les éléments transposables de classe 1 ou rétrotransposons et plus particulièrement les rétrotransposons à LTR. La recombinaison transpositionnelle est un processus au cours duquel un élément génétique mobile représenté par une séquence d'ADN est inséré dans un ADN cible. Cette insertion peut se produire selon deux mécanismes, directement à partir de l'ADN, ou par un intermédiaire à ARN. Les éléments mobiles qui sont transposés directement à partir de l'ADN sont appelés « transposons », et ceux qui sont transposés à partir d'un intermédiaire à ARN sont appelés « rétrotransposons ». Pour le transposon, ce dernier est excisé par une enzyme spéciale appelée « transposase » à activité nucléase et ligase et est directement inséré dans l'ADN cible. Pour le rétrotransposon, ce dernier est tout d'abord transcrit en ARN puis rétrotranscrit en ADN avant d'être inséré dans l'ADN cible (**figure 7**).

Les rétrovirus peuvent donc être considérés comme des éléments transposables puisque leur génome à ARN va être rétro-transcrit en ADN par la transcriptase inverse avant d'être intégré dans l'ADN cible par l'intégrase. Le provirus peut ainsi être considéré comme un rétrotransposon. Toutefois, une différence majeure entre les rétrovirus et les rétrotransposons réside dans le fait que ces derniers ont un cycle uniquement intracellulaire et ne peuvent pas être transmis à des cellules voisines par un mécanisme de réplication (Cox *et al.* 2011).

L'enzyme intégrase appartient à la classe des transposases DDE, caractérisées par la présence du motif DDE (triade catalytique composée de deux acides aspartiques et d'un acide glutamique) et d'un repliement du domaine catalytique de type Ribonucléase H ou RNaseH. Ces transposases sont retrouvées également chez certains bactériophages comme les bactériophages Mu, Tn5 et Tn10 (Rice & Baker 2001). Toutefois, les transposases appartenant à cette famille pourront présenter quelques nuances dans les mécanismes d'excision du transposon et d'insertion dans l'ADN cible.

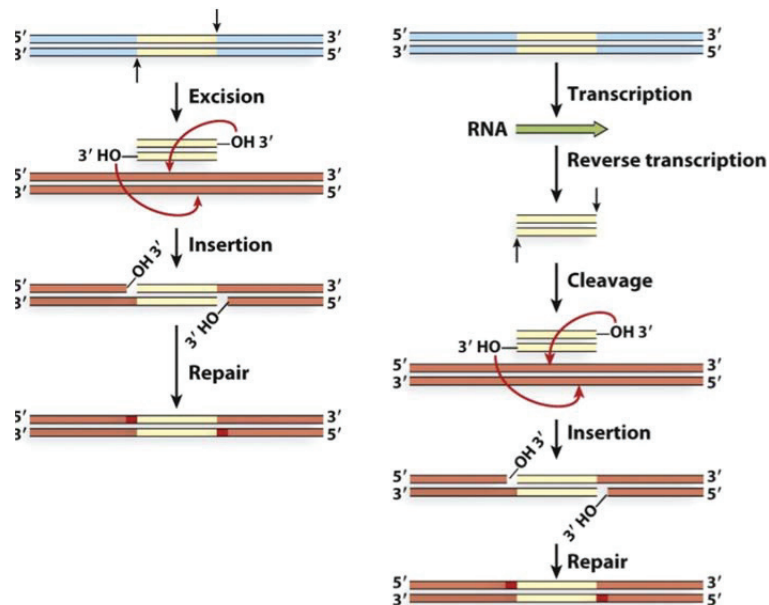


Figure 7 : Représentation schématique du mécanisme utilisé chez les transposons (**gauche**) et les rétrotransposons (**droite**) d'après (Cox *et al.* 2011).

3.2. Fonctions enzymatiques principales

L'intégration de l'ADN viral dans l'ADN de la cellule hôte est réalisée en trois étapes principales, dont deux sont catalysées directement par l'enzyme intégrase (Semenova *et al.* 2008) (**Figure 8**).

La première étape est l'étape dite de « clivage ». Les intégrases rétrovirales sont capables de catalyser cette réaction en présence de magnésium ou de manganèse. Cette étape consiste en un clivage des deux nucléotides terminaux des séquences inversées répétées (GT chez le VIH-1 et TT chez les ASLV) en 3' de chacun des brins de l'ADN viral bicaténaire rétro-transcrit. L'étape de clivage va permettre la création d'extrémités 3' hydroxylées se terminant par un dinucléotide CA, retrouvé chez l'ensemble des rétrovirus. Au cours de cette étape, l'intégrase va donc catalyser une attaque nucléophile d'une molécule d'eau sur le squelette phosphodiester de l'ADN viral avec la participation de deux cations métalliques divalents.

La deuxième étape est l'étape dite de « transfert de brins ». Cette réaction de transestérification nécessite également la présence de magnésium ou de manganèse comme pour l'étape de clivage. Cette étape consiste en une attaque nucléophile des groupements 3'OH des extrémités virales sur les liaisons phosphodiester de l'ADN hôte. Les brins de l'ADN cellulaire sont clivés avec un décalage de quatre à six paires de bases suivant le type viral (5 pb pour le VIH-1 et 6 pb

pour les ASLV). Cette réaction aboutit à la formation d'une structure intermédiaire où l'extrémité virale n'est liée à l'ADN cible que par un seul brin.

La troisième étape est l'étape de « remplissage » et n'est pas catalysée directement par l'intégrase. Dans cette étape, les deux nucléotides terminaux en 5' des deux brins de l'ADN viral sont clivés par une nucléase cellulaire et les brèches créées au niveau de l'ADN cellulaire sont comblées, de sorte qu'une duplication d'une courte séquence de l'ADN cellulaire soit générée de part et d'autre du provirus. Cette dernière étape n'est pas observée *in vitro*, des facteurs cellulaires de réparation de l'ADN seraient donc impliqués. A l'inverse, il a été montré que l'intégrase était nécessaire et suffisante aux deux premières étapes de clivage et de ligation (Katz *et al.* 1990).

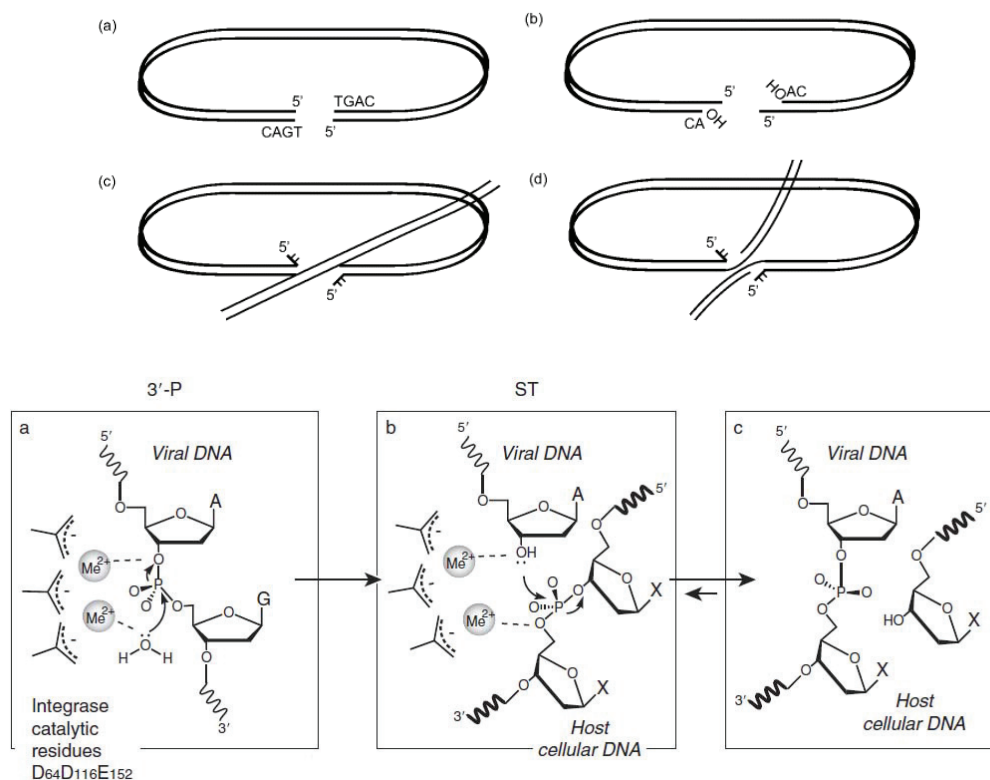


Figure 8 : (Haut) Intégration rétrovirale avec l'étape de clivage (b) et de transfert de brin (c) d'après (Neamati & Wang 2011).

(Bas) Mécanisme moléculaire de l'intégration rétrovirale (3'P = étape de clivage, ST = étape de transfert de brins) d'après (Semenova *et al.* 2008).

3.3. Fonctions enzymatiques secondaires

Sous certaines conditions, l'intégrase va être capable de catalyser certaines réactions dites « secondaires » qui contrairement aux réactions de clivage et de transfert de brins, ne semblent pas avoir de rôle dans le processus général d'intégration (Neamati & Wang 2011).

Lorsque l'intégrase est incubée avec un ADN mimant le produit de la réaction de transfert de brins et en présence de cations métalliques divalents, une réaction dite de « désintégration » qui correspond à l'inverse de la réaction de transfert de brins peut se produire. Cette réaction va aboutir à la libération de l'oligonucléotide mimant l'extrémité virale et permet de reformer la liaison phosphodiester au niveau de l'ADN cible (Chow *et al.* 1992). Cette réaction n'a jamais été observée *in*

vivo et serait probablement néfaste pour la réplication du virus. Contrairement aux étapes de clivage et de transfert de brins qui nécessitent la protéine entière, la réaction de désintégration ne nécessite que le domaine catalytique de l'intégrase.

Sous certaines conditions, l'intégrase peut également présenter une activité nucléosidase non spécifique. Cette activité fut même la première détectée chez une intégrase rétrovirale (Grandgenett *et al.* 1978). Cette réaction est permise par l'intégrase qui va catalyser l'attaque nucléophile d'un oxygène d'une molécule d'eau sur la liaison phosphodiester de l'ADN cible. En addition de cette réaction d'hydrolyse, l'intégrase peut aussi cliver l'ADN par une alcoololyse non spécifique (Katzman & Sudol 1996). Pour cette dernière réaction, l'élément nucléophile est constitué par le groupement hydroxyle d'un alcool inclut dans le milieu réactionnel (comme le glycérol). Dans le produit de cette réaction, l'alcool va être lié à l'extrémité 5' phosphate de l'ADN au niveau du site de clivage. Les alcools pourront d'ailleurs également se substituer à la molécule d'eau nucléophile au cours de la première étape de l'intégration ou étape de clivage. Cette réaction n'a également jamais été observée *in vivo*.

3.4. Fonctions non enzymatiques

Les mutants de l'intégrase ont été historiquement répartis en deux classes (Engelman 1999). La première classe regroupe l'ensemble des mutants dits « de classe 1 » pour lesquels l'intégrase est inactive pour sa fonction principale, qui est la fonction d'intégration. Il a cependant pu être observé que de nombreuses mutations de l'intégrase pouvaient avoir des conséquences sur d'autres étapes du cycle réplcatif rétroviral. Ces mutants dits « de classe 2 » mettent en évidence que les fonctions de l'intégrase ne se limitent pas à l'étape d'intégration et que cette enzyme intervient dans d'autres étapes de la réplication rétrovirale.

Les effets de certains mutants de l'intégrase sur d'autres étapes du cycle rétroviral ont pu être décrits pour les étapes de décapsidation (Briones *et al.* 2010), de transcription inverse (Tsurutani *et al.* 2000), d'import nucléaire de l'ADN viral rétro-transcrit (Ikeda *et al.* 2004) et de maturation des virions (Mohammed *et al.* 2011).

L'ensemble des mutations de classe 2 ne concerne pas systématiquement les résidus catalytiques du motif DDE. Les effets pléiotropiques de ces mutations ne seraient donc pas toujours liés directement à la perte de son activité enzymatique (Masuda 2011).

4. Le complexe de pré-intégration

De nombreuses études ont mis en évidence que le processus d'intégration *in vivo* ne faisait pas intervenir uniquement la protéine intégrase mais également des protéines cellulaires de l'hôte, ainsi que d'autres protéines virales. L'association *in vivo* de l'intégrase avec ses partenaires protéiques viraux et cellulaires constitue le complexe de pré-intégration ou PIC (pour *Pre-Integration Complex*) (Farnet & Haseltine 1990). Chez le VIH-1, ce complexe de pré-intégration contiendrait l'ADN viral rétrotranscrit sous forme linéaire ainsi que différents partenaires protéiques viraux dont la protéine de matrice, la transcriptase inverse, l'intégrase, la protéine de nucléocapside et la protéine de capsid. Différents partenaires cellulaires ont également été identifiés dont les protéines LEDGF/p75 (pour *Lens Epithelium-Derived Growth Factor*), INI1 (pour *Integrase Interactor 1*), BAF (pour *Barrier to Autointegration Factor*) et HMGA1 (pour *High-Mobility Group proteins*) (Krishnan & Engelman 2012). Les partenaires cellulaires et viraux du PIC vont pouvoir réguler le processus d'intégration à différents niveaux : au niveau de la catalyse enzymatique, du transport du PIC vers le

compartiment nucléaire, de la sélection des sites d'intégration dans l'ADN cible ou encore de la réparation des brèches de l'ADN viral suite à l'étape de transfert de brins.

L'étape de clivage de l'ADN viral est réalisée dans le cytoplasme, alors que l'étape de transfert de brins est réalisée dans le noyau. Il a donc été proposé que la fonction principale du PIC était de séparer dans le temps ces deux réactions *in vivo* pour prévenir le phénomène d'auto-intégration, où l'ADN viral est utilisé comme ADN cible pour l'étape de transfert de brins, ce qui est fréquemment observé *in vitro* (Semenova *et al.* 2008). L'intégrase pourrait au sein du PIC, être maintenue sous une forme inactive dans le cytoplasme jusqu'à son passage dans le noyau. Il a en effet été montré que la protéine cellulaire BAF était présente au sein du PIC du VIH-1 et permettait d'empêcher le phénomène d'auto-intégration (Zheng *et al.* 2000). Un autre partenaire cellulaire nommé « INI1 », semble empêcher les activités de clivage et de transfert de brins de l'intégrase du VIH-1 (Maillot *et al.* 2013). D'autres partenaires de l'intégrase au sein du PIC pourraient présenter cette même fonction comme la transcriptase inverse virale qui est capable d'inhiber l'activité catalytique de l'intégrase *in vitro* (Oz *et al.* 2002). L'intégrase pourrait être réactivée suite à son transfert au noyau par l'intermédiaire de protéines cellulaires nucléaires ou de l'interaction du PIC avec la chromatine. A l'inverse, d'autres protéines du PIC semblent capables d'augmenter l'activité catalytique de l'intégrase. Chez le VIH-1, ce phénomène a été observé pour la protéine virale de nucléocapside mais également pour la protéine cellulaire LEDGF/p75 (Carteau *et al.* 1999; Cherepanov 2007).

Les éléments du PIC sont également impliqués dans l'adressage de ce dernier au noyau, en particulier pour le passage du PIC par le biais des pores nucléaires lors de l'infection de cellules quiescentes (cas des lentivirus). Pour les virus infectant les cellules quiescentes, le PIC entrerait dans le nucléoplasme par l'intermédiaire de pores nucléaires. La taille du PIC a un diamètre bien supérieur à la taille du canal central des pores nucléaires. L'entrée dans le noyau nécessite donc un transport actif. Des protéines virales et cellulaires peuvent contribuer au transport du PIC à travers la membrane nucléaire favorisant ainsi l'accès à la chromatine et assurant l'intégration dans des domaines particuliers de celle-ci. Des études ont permis d'identifier des signaux de localisation nucléaire (NLS pour *Nuclear Localization Sequence*) portés par trois protéines virales différentes : IN, MA et Vpr. Ces NLS sont reconnues par des protéines cellulaires appelées importines et permettent un transport actif de molécules du cytoplasme vers le noyau. L'association de ces trois séquences contribuerait à un ciblage nucléaire réussi (Suzuki & Craigie 2007). Il a été également montré que la présence d'une structure nommée « DNA Flap central » était un facteur contribuant à l'import nucléaire du PIC. Cette structure d'ADN à trois brins, présente au centre du génome du VIH-1, est formée lors de la transcription inverse. Un ADN viral sans la séquence DNA flap se retrouve accumulé au voisinage de la membrane nucléaire (Zennou *et al.* 2000). Le PIC pourrait aussi interagir avec un facteur nommé « TNPO3 » (pour *TrANsPORtin-3*) pour son transport dans le noyau. TNPO3 est une protéine de la famille des importines β et elle a été retrouvée en interaction avec l'IN d'une part, et avec le complexe CA-NC du virus VIH-1, d'autre part. TNPO3 est requise pour qu'il y ait infection par des virus du genre *Lentivirus*. En effet, une déplétion de TNPO3 inhibe la réplication du virus VIH-1. Cette inhibition se fait après l'étape d'import nucléaire mais avant l'étape d'intégration (Cribier *et al.* 2011; Valle-Casuso *et al.* 2012).

Des zones d'intégration préférentielles dans le génome ont pu être définies pour certains rétrovirus. Il a pu être montré que les gammarétrovirus ainsi que les spumavirus s'intégraient préférentiellement à proximité de sites d'initiation de la transcription ou d'îlots CpG (Moalic *et al.* 2006; Tsukahara *et al.* 2006). Chez le VIH-1 (lentivirus), l'intégration aurait préférentiellement lieu dans des régions fortement transcrites à l'intérieur même des gènes (Lewinski *et al.* 2006). Les alpharétrovirus et les deltarétrovirus ont une faible préférence d'intégration dans les unités de transcription ou les îlots CpG (Narezkina *et al.* 2004). Les bétarétrovirus ne montrent quant à eux pas de préférence dans le choix du site d'intégration (Faschinger *et al.* 2008). Ce ciblage préférentiel de

l'intégrase vers certaines régions de l'ADN génomique pourrait faire intervenir des constituants du PIC. Chez le VIH-1, il a été montré que l'ADN viral interagissait avec la chromatine par l'intermédiaire d'une protéine nucléaire appelée Emerin qui est associée au PIC par l'intermédiaire de BAF. Il a ainsi été montré que Emerin et BAF étaient nécessaires au ciblage de l'intégrase vers l'ADN chromosomique (Jacque & Stevenson 2006). Un autre partenaire de l'intégrase du VIH-1, LEDGF/p75, est essentiel pour le ciblage de l'intégrase à l'ADN chromosomique et en particulier dans les régions activement transcrites (Engelman & Cherepanov 2008). Il a ainsi pu être montré que des cellules n'exprimant pas LEDGF/p75 ne permettaient pas une réplication efficace des virions (Maertens *et al.* 2003; Emiliani *et al.* 2005). L'interaction de l'intégrase du VIH-1 avec la protéine INI1 faisant partie du complexe SWI/SNF (pour *SWItch/Sucrose NonFermentable*), intervenant dans le remodelage de la chromatine, pourrait également faciliter l'adressage de l'intégrase à l'ADN cible ainsi que les protéines HMG (Turelli *et al.* 2001; Gao *et al.* 2003; Lesbats *et al.* 2011).

L'intégration stable du provirus requiert, après intégration, la réparation des espaces nucléotidiques encadrant l'élément intégré. Plusieurs protéines cellulaires appartenant au PIC ont été proposées comme candidates à ce rôle. La plupart de ces protéines a des fonctions de réparation de l'ADN. Il a ainsi été reporté une interaction entre l'intégrase du VIH-1 et des facteurs de la famille RAD impliquée dans la réparation des cassures double brin de l'ADN (Mulder *et al.* 2002).

5. Propriétés structurales

5.1. Structures primaires

Les intégrases rétrovirales comprennent généralement près de 290 résidus et sont constituées de trois domaines, amino-terminal (résidus 1 à 52 chez le RSV et 1 à 55 chez le VIH-1), central ou catalytique (résidus 52 à 207 chez le RSV et 55 à 212 chez le VIH-1) et carboxy-terminal (résidus 207 à 286 chez le RSV et 212 à 270 chez le VIH-1) qui sont reliés entre eux par des boucles flexibles (Jaskolski *et al.* 2009) (**figure 9**). Il a été montré que ces trois domaines étaient nécessaires à la formation d'un complexe stable entre l'intégrase et l'ADN viral, et à la réalisation des étapes de clivage et de ligation (Vink *et al.* 1994). Certaines variations sont à noter comme pour l'intégrase du prototype humain des spumavirus ou PFV (pour *Prototype Foamy Virus*), qui comprend 392 résidus et l'intégrase des ASLV qui comprend 323 résidus avant de subir un clivage post-traductionnel aboutissant à une protéine active de 286 résidus (Katz & Skalka 1988).

Les intégrases de différentes souches rétrovirales possèdent des identités et similarités de séquences variables. Pour exemple, l'intégrase du VIH-1 (intégrase lentivirale) présente 57/74 et 23/37 % d'identité/similarité de séquence avec l'intégrase du SIV (pour *Simian Immunodeficiency Virus*, intégrase lentivirale) et d'ASV (intégrase non lentivirale) respectivement. Les identités et similarités de séquence sont plus élevées sur les domaines seuls, et les structures isolées de ces domaines présentent des structures tridimensionnelles très proches. De façon générale, il est observé que la conservation de séquence des intégrases rétrovirales est la plus faible au niveau du domaine carboxy-terminal (Jaskolski *et al.* 2009).

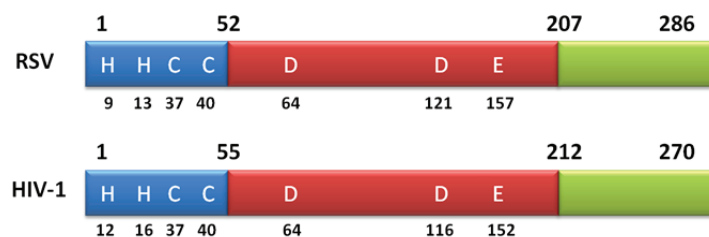


Figure 9 : Structures primaires des intégrases du RSV et du VIH-1. Le domaine N-terminal est en bleu, le domaine catalytique en rouge et le domaine C-terminal en vert. La triade catalytique DDE (domaine central) et le motif HHCC (domaine amino-terminal) sont représentés.

5.2. Structures tridimensionnelles

5.2.1. Le domaine amino-terminal ou NTD

La structure du domaine amino-terminal ou NTD (pour *N-Terminal Domain*) a été déterminée chez le VIH-1 (Cai *et al.* 1997) et le VIH-2 (Eijkelenboom *et al.* 1997) par RMN (**figure 10**). Il est constitué de quatre hélices α avec un motif hélice-tour-hélice caractéristique des protéines se fixant aux acides nucléiques et se présente sous forme dimérique. Il possède un repliement semblable aux répresseurs bactériens hélice-tour-hélice, qui est conservé dans toutes les intégrases rétrovirales et de rétrotransposons.

La structure cristalline de la protéine intégrase du VIH-1, contenant les domaines N-terminaux et centraux, fut ensuite résolue et permis de définir une interface de dimérisation du domaine amino-terminal différente de celle du domaine amino-terminal isolé (Wang *et al.* 2001). L'interaction entre les deux domaines amino-terminaux semble donc flexible et peut adopter différentes conformations (**figure 11**).

La structure cristalline de l'intégrase du VIH-2, contenant les domaines amino-terminaux et centraux et complexée avec le partenaire cellulaire LEDGF par son *Integrase Binding Domain* ou IBD, a également été résolue (Hare *et al.* 2009) (**figure 11**). Dans cette structure, le domaine amino-terminal ne se dimérise pas, mais interagit avec le domaine central et le partenaire cellulaire LEDGF. Un résultat similaire a été observé pour la structure cristalline de l'intégrase du *Maedi Visna Virus* ou MMV contenant les domaines amino-terminaux et centraux avec le partenaire cellulaire LEDGF (Hare *et al.* 2009). Dernièrement, la structure atomique de l'intégrase entière du PFV en présence d'ADN a montré un arrangement spatial du NTD par rapport aux autres domaines encore différent, dans lequel le NTD interagit avec l'ADN et le CCD du monomère opposé. Le mode d'interaction du NTD n'est donc pas définitivement résolu (Hare *et al.* 2010).

Le domaine amino-terminal comporte quatre résidus très conservés qui sont deux histidines (9 et 13 pour les ASLV et 13 et 16 pour le VIH-1) et deux cystéines (37 et 40 pour les ASLV et 40 et 43 pour le VIH-1). Ce motif appelé HHCC est capable de fixer un ion zinc mais ne présente pas une structure similaire aux structures en doigt à zinc classiques (Eijkelenboom *et al.* 1997). Des mutations dans le motif HHCC ou une délétion de ce domaine diminuent l'habilité de l'intégrase à catalyser l'étape de clivage et de transfert de brin avec maintien de l'activité de désintégration (Bushman & Wang 1994). Le domaine N-terminal n'a donc aucun rôle direct dans la catalyse mais pourrait jouer un rôle dans la reconnaissance et le positionnement des extrémités virales au cours du processus intégratif (Yang & Roth 2001). Le domaine amino-terminal interagit également avec l'ADN cellulaire (Heuer & Brown 1997). Il est également impliqué dans la multimérisation de l'enzyme entière. La fixation d'ions zinc au niveau de ce domaine permet de stabiliser la structure protéique. Elle engendre la formation d'une structure plus ordonnée avec augmentation du nombre d'hélices α et

favorise l'oligomérisation, notamment la formation de tétramères d'intégrase (Zheng *et al.* 1996). Chez le VIH-1, le domaine amino-terminal est la région préférentielle de liaison pour le facteur transcriptionnel cellulaire LEDGF (Cherepanov *et al.* 2003).

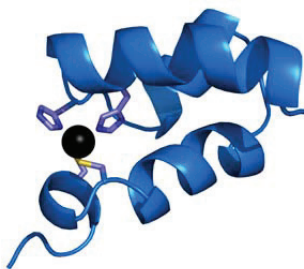


Figure 10 : Structure RMN du NTD de l'intégrase du VIH-1 (NTD en bleu, zinc en noir, résidus HHCC en cyan, PDB 1WJC).

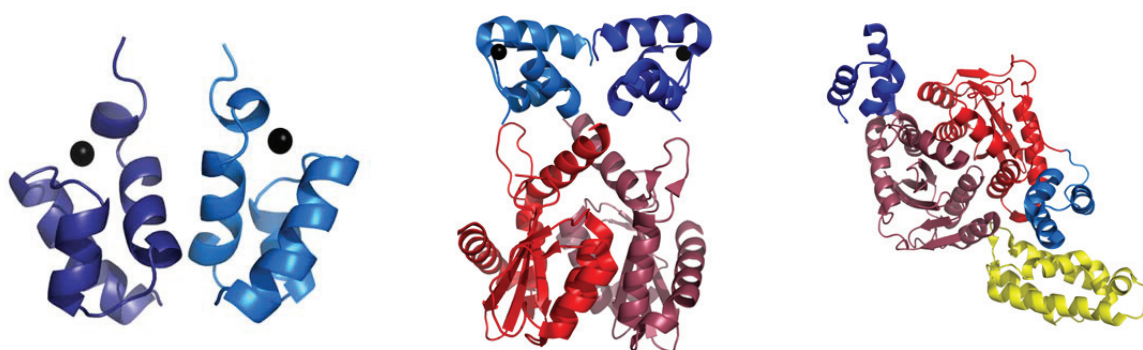


Figure 11 : **(Gauche)** Structure RMN du NTD sous forme dimérique de l'intégrase du VIH-1 (NTD en bleu et zinc en noir, PDB 1WJC). **(Milieu)** Structure cristalline de la protéine IN NTD+CCD sous forme dimérique du VIH-1 (NTD en bleu et CCD en rouge, PDB 1K6Y). **(Droite)** Structure cristalline de la protéine IN NTD+CCD sous forme dimérique du VIH-2 en complexe avec IBD-LEDGF (NTD en bleu, CCD en rouge, IBD-LEDGF en jaune, PDB 3F9K).

5.2.2. Le domaine central catalytique ou CCD

Une trentaine de structures cristallines du CCD (pour *Catalytic Central Domain*) a été obtenue chez différentes intégrases rétrovirales dont le VIH-1 (Dyda *et al.* 1994; Bujacz *et al.* 1996; Goldgur *et al.* 1998; Maignan *et al.* 1998; Goldgur *et al.* 1999; Greenwald *et al.* 1999; Chen *et al.* 2000; Molteni *et al.* 2001), ASV (Bujacz *et al.* 1995; Bujacz *et al.* 1996; Bujacz *et al.* 1997; Lubkowski *et al.* 1998; Lubkowski *et al.* 1998; Lubkowski *et al.* 1999), PFV (Valkov *et al.* 2009), BIV (Yao *et al.* 2010) et PrimateFV (Rety *et al.* 2010). Un modèle RMN a également été obtenue pour le domaine catalytique de l'intégrase du VIH-1 et confirme la structure cristalline précédemment déterminée (Fitzkee *et al.* 2010).

Ce domaine se présente sous la forme d'un dimère (**figure 12**). La conformation tridimensionnelle du domaine central est de type RnaseH et est très conservée parmi les membres de la famille des phosphotransférases DDE, qui inclut les intégrases des rétrovirus et les rétrotransposases. Le cœur catalytique est composé d'un mélange d'hélices α et de brins β . Pour le RSV, ce domaine est constitué de cinq brins β et de cinq hélices α et pour le VIH-1 de cinq brins β et de six hélices α . La conformation tridimensionnelle de ce domaine montre que les trois résidus très conservés et porteurs de l'activité catalytique sont très proches.

Le domaine catalytique de l'intégrase ou CCD contient les résidus catalytiques de la triade DDE, mais présente une activité enzymatique limitée en l'absence des deux autres domaines de l'enzyme. Le domaine catalytique isolé n'est pas capable de catalyser la réaction de transfert de brins, cependant, il est capable de catalyser la réaction de clivage avec toutefois une spécificité considérablement réduite (Kulkosky *et al.* 1995). Le domaine catalytique est néanmoins capable de catalyser la réaction de désintégration en l'absence des domaines amino et carboxy-terminaux. Ce domaine est également impliqué dans la reconnaissance des extrémités virales (Agapkina *et al.* 2006). Chez le VIH-1, le domaine central est impliqué dans la liaison avec le facteur transcriptionnel cellulaire LEDGF (Cherepanov *et al.* 2005).



Figure 12 : Structure cristalline du CCD sous forme dimérique de l'intégrase du VIH-1 (triade catalytique en jaune, magnésium en vert, PDB 1BL3).

5.2.3. Le domaine carboxy-terminal ou CTD

Le domaine carboxy-terminal ou CTD (pour *C-Terminal Domain*) est le domaine le moins conservé entre les intégrases des différents rétrovirus. La structure tridimensionnelle de ce domaine pour l'intégrase du VIH-1 a été résolue par RMN par deux groupes (Eijkelenboom *et al.* 1995; Lodi *et al.* 1995; Eijkelenboom *et al.* 1999). Il comporte cinq brins β antiparallèles arrangés en un tonneau β (**figure 13**). Il présente une structure similaire à un domaine SH3 (domaine retrouvé dans de nombreuses protéines qui interagissent avec d'autres protéines ou avec les acides nucléiques) et se présente sous forme de dimère.

Les structures de la protéine intégrase du VIH-1 et du RSV contenant les domaines centraux et carboxy-terminaux ont également été résolues par cristallographie aux rayons X (Chen *et al.* 2000; Yang *et al.* 2000). Pour le VIH-1, seuls les domaines centraux interagissent entre eux alors que les deux domaines carboxy-terminaux se présentent sous forme de monomère de part et d'autre des domaines centraux. Pour le RSV, les domaines carboxy-terminaux tout comme les domaines centraux interagissent entre eux avec une interface de dimérisation du domaine carboxy-terminal différente de celle observée pour le domaine isolé chez le VIH-1 (**figure 14**).

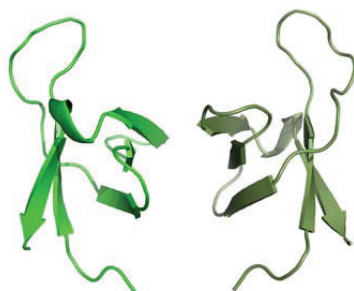


Figure 13 : Structure RMN du CTD sous forme dimérique de l'intégrase du VIH-1 (PDB 1IHV).

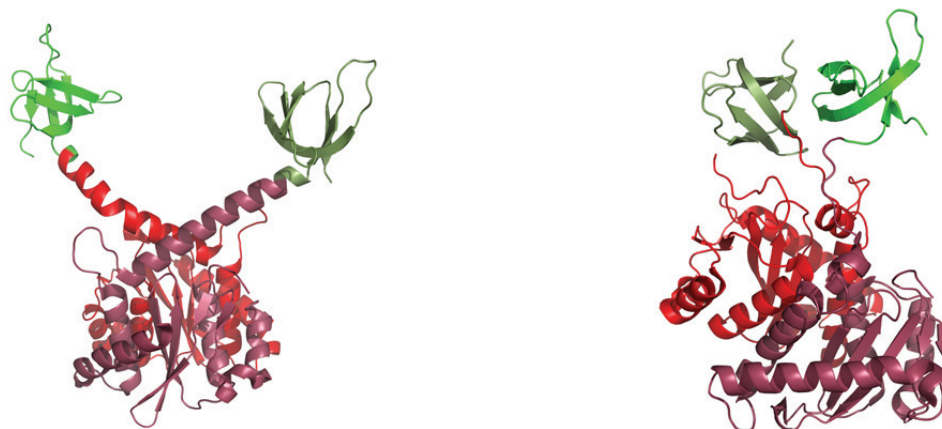


Figure 14 : **(Gauche)** Structure cristalline de la protéine IN CCD+CTD sous forme dimérique du VIH-1 (CTD en vert et CCD en rouge, PDB 1EX4).

(Droite) Structure cristalline de la protéine IN CCD+CTD sous forme dimérique de l'intégrase du RSV (CTD en vert et CCD en rouge, PDB 1C0M).

Le CTD est impliqué dans la fixation à l'ADN hôte. En effet, la suppression de ce domaine diminue fortement la capacité de la protéine à se fixer sur l'ADN (Lutzke & Plasterk 1998). Ce domaine est aussi impliqué dans la reconnaissance et la fixation de l'ADN viral. Des expériences de pontage chimique réalisées chez le VIH-1 ont montré que les extrémités virales interagissaient avec le domaine C-terminal de la protéine intégrase (Esposito & Craigie 1998). Des mutations réalisées dans ce domaine inhibent l'oligomérisation de la protéine intégrase (Kalpana *et al.* 1999). Le domaine carboxy-terminal a également un rôle dans le bon repliement de la protéine entière (Moreau *et al.* 2003). Le CTD de l'intégrase du VIH-1 lie la protéine cellulaire EED (pour *Embryonic Ectoderm Development*) et cette liaison augmente son activité d'intégration *in vitro* (Violot *et al.* 2003). Ce domaine est également capable d'interagir avec la transcriptase inverse et cette interaction serait nécessaire à la réaction de transcription inverse (Zhu *et al.* 2004).

6. Etat oligomérique de la protéine entière

Des analyses par ultracentrifugation analytique et chromatographie d'exclusion de taille sur l'intégrase et ses domaines ont permis de déterminer que l'intégrase pouvait se trouver sous forme monomérique ou s'auto-assembler *in vitro* pour former des dimères, des tétramères et parfois des octamères. Il a été montré que de nombreux éléments pouvaient influencer la formation de ces oligomères tels que la concentration protéique, la présence d'ADN, le pourcentage de glycérol, la concentration saline du tampon utilisé, la présence de détergents, la température ou encore la présence de cations divalents comme le magnésium, le manganèse ou encore le zinc (Jones *et al.* 1992; Zheng *et al.* 1996; Lee *et al.* 1997; Coleman *et al.* 1999; Deprez *et al.* 2000).

Des études ont montré que la multimérisation était nécessaire à l'intégrase pour les activités de clivage et de ligation. Il a été proposé que la protéine serait active sous forme de dimère lors de l'étape de clivage et sous forme de tétramère lors de l'étape de transfert de brins (Faure *et al.* 2005).

Dans le cadre du processus intégratif, plusieurs modèles d'oligomérisation ont été proposés : le modèle dimérique (Yang *et al.* 2000), le modèle tétramérique (Wang *et al.* 2001; Wielens *et al.* 2005; Dolan *et al.* 2009) ou encore le modèle octamérique (Heuer & Brown 1998). Pour les intégrases du VIH-1 et du RSV, l'espace compris entre les deux sites actifs de la structure dimérique ne correspond pas avec l'espace existant entre les deux sites de clivage au niveau de l'ADN cible. En effet, les deux sites actifs de la forme dimérique sont séparés par 45 Å alors que les cinq ou six paires de bases séparant les deux sites de coupure au niveau de l'ADN cible correspondent à un espace de 16 à 20 Å. A l'inverse, le modèle tétramérique permet de justifier l'espacement existant entre les deux sites de coupure au niveau de l'ADN cible. Des observations directes par microscopie à force atomique de l'intégrase complexée avec de l'ADN ont montré que ces complexes avaient la taille apparente d'un tétramère (Bao *et al.* 2003). Des résultats similaires ont été obtenus par cryo-microscopie électronique avec un modèle obtenu à 27 Å de résolution (Ren *et al.* 2007). Des tétramères d'intégrase ont également pu être isolés d'extraits nucléaires, en particulier un tétramère d'intégrase lié à l'ADN et complexé à LEDGF (Cherepanov *et al.* 2003).

Cette possibilité d'un modèle tétramérique pour l'étape d'intégration a récemment été renforcée par le modèle obtenu par cryo-microscopie électronique de l'intégrase du VIH-1 avec son partenaire cellulaire LEDGF, révélant une intégrase sous forme tétramérique, mais également par la première structure cristalline de l'intégrase entière du PFV complexée avec l'ADN viral et l'ADN cible impliquant une forme tétramérique de l'intégrase (Michel *et al.* 2009; Hare *et al.* 2010; Maertens *et al.* 2010) (**figure 15**). L'observation d'une réaction de clivage et de transfert de brins *in cristallo* sur la forme tétramérique de l'intégrase entière du PFV suggère également la possibilité que l'étape de clivage soit, comme pour l'étape de transfert de brins, catalysée par un tétramère d'intégrase (Hare *et al.* 2012). Un effet inhibiteur du partenaire cellulaire IN1 sur l'activité de clivage de l'intégrase du VIH-1 a également été observé et semble impliquer une interaction de ce partenaire avec une intégrase sous forme tétramérique soulignant là aussi la possibilité d'une forme tétramérique pour l'activité de clivage (Maillot *et al.* 2013).

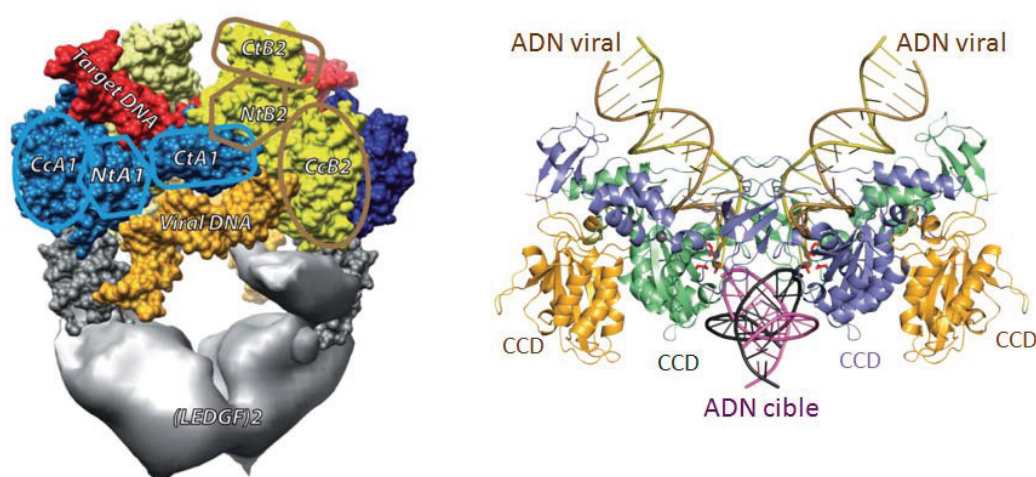


Figure 15 : (**Gauche**) Reconstruction 3D par cryo-microscopie électronique d'un tétramère d'intégrase du VIH-1 avec l'ADN viral (orange), l'ADN cellulaire (rouge) et le cofacteur LEDGF (gris) d'après (Michel *et al.* 2009).

(**Droite**) Structure cristalline de l'intasome du prototype humain des spumavirus (PFV) : tétramère d'intégrase avec l'ADN viral et cellulaire (cible) d'après (Maertens *et al.* 2010).

Dans les deux modèles présentés ci-dessus, le tétramère semble être un dimère de dimère. L'arrangement des différents domaines de l'intégrase est différent pour les deux modèles, avec un arrangement au sein des protomères de type CCD-CTD-NTD pour le PFV et de type CCD-NTD-CTD pour le VIH-1. De plus, l'assemblage du tétramère pour le modèle du VIH-1 semble moins compact que pour le modèle de PFV, où l'intégrase et l'ADN sont en interactions très étroites. Le modèle du VIH-1 met en évidence l'importance du cofacteur cellulaire LEDGF qui sert de plateforme à la tétramérisation de l'enzyme. Sur le modèle PFV, l'absence de cofacteur cellulaire semble impliquer un mode d'oligomérisation différent qui fait intervenir l'ADN comme support de la multimérisation. Il est à remarquer que les domaines carboxy et amino-terminaux d'un des monomères de chaque dimère n'est pas visible dans le modèle PFV du fait de l'agitation de ces derniers. Même si des études SAXS récentes ont permis de déterminer leurs localisations potentielles au sein du tétramère (Gupta *et al.* 2012), leurs rôles dans le processus d'intégration reste à déterminer.

D'autres résultats récents ont également permis d'obtenir un modèle SAXS de l'intégrase entière du RSV appuyé par des expériences de pontage chimique (Bojja *et al.* 2011). Ces résultats suggèrent que l'intégrase serait dimérique en solution en absence d'ADN et que la structure du dimère ressemble au dimère interne observé sur la structure PFV avec un rôle important du CTD dans la dimérisation. De façon intéressante, l'interaction CCD-CCD ne semble pas intervenir sur le dimère, mais uniquement sur le tétramère. Plus récemment, la même équipe a réalisé une étude SAXS sur l'intégrase du VIH-1 sous forme tétramérique et sous forme dimérique (mutant), qui serait capable d'adopter deux conformations dimériques possibles impliquant soit une interaction CCD-CCD en absence de zinc, soit une interaction de type « dimère interne PFV » en présence de zinc appuyant le rôle du domaine amino-terminal dans la multimérisation (Bojja *et al.* 2013). Le passage de ses deux formes dimériques ainsi que l'intervention de formes monomériques dans l'assemblage de la forme tétramérique restent mal connus.

7. Les stratégies anti-intégrase

L'intégration de l'ADN viral au sein du génome cellulaire est nécessaire à la formation de nouvelles particules virales et à la persistance de l'infection. La protéine intégrase est donc une cible idéale pour lutter contre la propagation des virus. Elle est également la seule des trois enzymes virales pour laquelle il n'existe actuellement qu'un inhibiteur utilisable en thérapie (Raltégravir). En effet, des inhibiteurs des deux autres enzymes, la transcriptase inverse et la protéase, sont déjà utilisés dans le cadre de la multithérapie anti-VIH. L'émergence de souches du VIH-1 résistantes aux inhibiteurs actuellement disponibles a mis à jour la nécessité de développer de nouveaux inhibiteurs spécifiques de cette protéine (Demeulemeester *et al.* 2012).

Plusieurs stratégies peuvent être envisagées pour inhiber l'intégrase.

Une des stratégies consiste à perturber les interactions entre l'intégrase et ses substrats ADN. Par exemple des oligonucléotides modifiés mimant la structure de la séquence *att* vont pouvoir dissocier les complexes intégrase-ADN viral. Pour cela, l'oligonucléotide utilisé doit avoir une affinité très importante pour la protéine (Brodin *et al.* 2002). Par ailleurs, des peptides peuvent être modifiés afin de pouvoir interagir avec la protéine intégrase et couvrir son domaine d'interaction avec l'ADN viral : des inhibiteurs qui correspondent à des fragments du domaine central de l'intégrase du VIH-1 ont déjà été développés (Maroun *et al.* 1999; De Soultrait *et al.* 2002). Ces deux peptides interagissent avec l'intégrase au niveau de son site actif et empêchent la fixation de l'enzyme sur son substrat ADN viral. Les inhibiteurs de la fixation de l'ADN viral à l'intégrase vont inhiber aussi bien l'activité de clivage que l'activité de transfert de brins.

D'autres inhibiteurs ont également été développés et visent à empêcher l'interaction du complexe intégrase-ADN viral avec l'ADN cellulaire après l'étape de clivage. On parle plus spécifiquement d'inhibiteurs du transfert de brins. En effet, ces composés ne présentent pas ou peu d'effets sur l'activité de clivage. Ceux-ci sont le plus souvent des dérivés d'acides dicétoniques et vont cibler la triade catalytique DDE et chélater les cations divalents du site actif. De plus, ils nécessitent la présence de l'ADN viral associé à l'enzyme à la suite de l'étape de clivage. L'interaction de ces composés avec ce complexe va empêcher la fixation de l'ADN cible et donc l'étape de transfert de brins. Le Raltégravir fait parti de ces composés et a été observé dans le site actif de l'intégrase de PFV en interaction étroite avec l'ADN viral clivé (par l'intermédiaire d'un groupement benzyle) et deux cations manganèse (par l'intermédiaire de la fonction dicétone) (**figure 16**) (Hare *et al.* 2010).

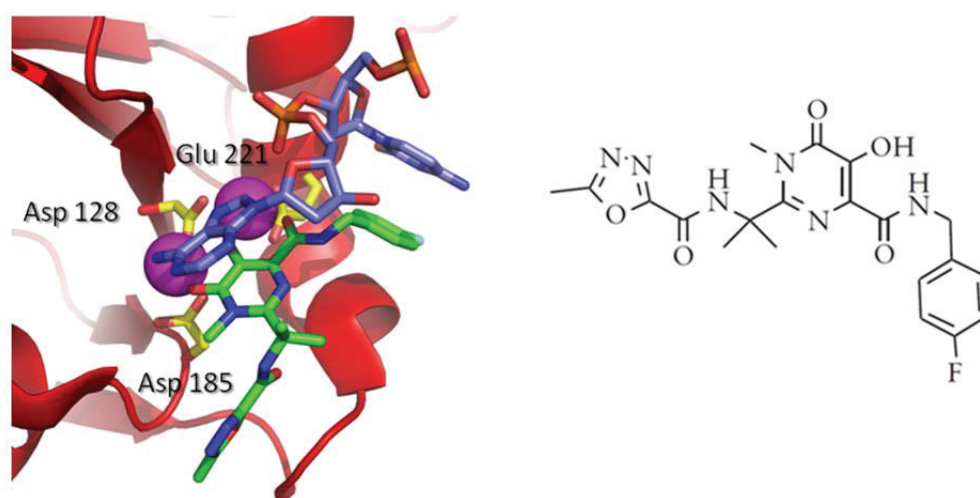


Figure 16 : (**Gauche**) Site actif de l'intasome du PFV complexé avec l'inhibiteur Raltégravir (résidus catalytiques en jaune, manganèses en magenta, Raltégravir en vert, extrémité CA 3' de l'ADN viral en bleu, CCD en rouge, PDB 3L2V). (**Droite**) Molécule de Raltégravir.

En dehors des inhibiteurs du site actif pour lesquels des résistances émergent très rapidement, une stratégie alternative vise à développer des inhibiteurs dits « allostériques » et qui vont pouvoir cibler d'autres endroits de l'enzyme que le site actif.

L'intégrase est active sous forme d'oligomères et la modulation de l'oligomérisation constitue une option intéressante pour son inhibition.

La plupart de ces études ont ainsi ciblé l'interface de dimérisation du domaine catalytique CCD. En effet, celle-ci est observée sur l'ensemble des structures cristallines d'intégrases dimériques connues à ce jour. Plusieurs de ces études ont reporté l'inhibition de l'intégrase par l'utilisation de peptides dérivés de séquences d'acides aminés impliqués dans la dimérisation du CCD (Maroun *et al.* 2001; Li *et al.* 2006; Zhao & Chmielewski 2013). Le mode d'action de ces peptides a été défini, ces derniers inhibant la dimérisation de l'enzyme par un mécanisme de compétition. D'autres études basées sur l'utilisation de composés chimiques ont également été réalisées. Parmi ces études, on peut citer la découverte de dérivés coumariniques qui se lient au niveau du CCD du VIH-1 en un site différent du site actif. La fixation de ces composés entraîne un réarrangement de certains résidus de la protéine (situés à l'interface de dimérisation du CCD), induisant une gêne stérique pour la formation du dimère, et inhibant ainsi les activités de l'enzyme (Al-Mawsawi *et al.* 2006).

La récente observation d'une molécule de sucrose dans un sillon et en un autre site au sein de l'interface de dimérisation du CCD de l'intégrase du VIH-1, constitue une piste intéressante pour

le développement de nouveaux inhibiteurs allostériques (**figure 17B**) (Wielens *et al.* 2010). Bien que ces auteurs n'aient pas observé d'effet inhibiteur du sucrose sur les activités de l'intégrase, d'autres études ont pu mettre en évidence la possible fixation de composés au niveau de ce sillon avec pour conséquence l'inhibition des activités catalytiques de l'enzyme par un mécanisme encore mal déterminé (Du *et al.* 2008; Kessl *et al.* 2009).

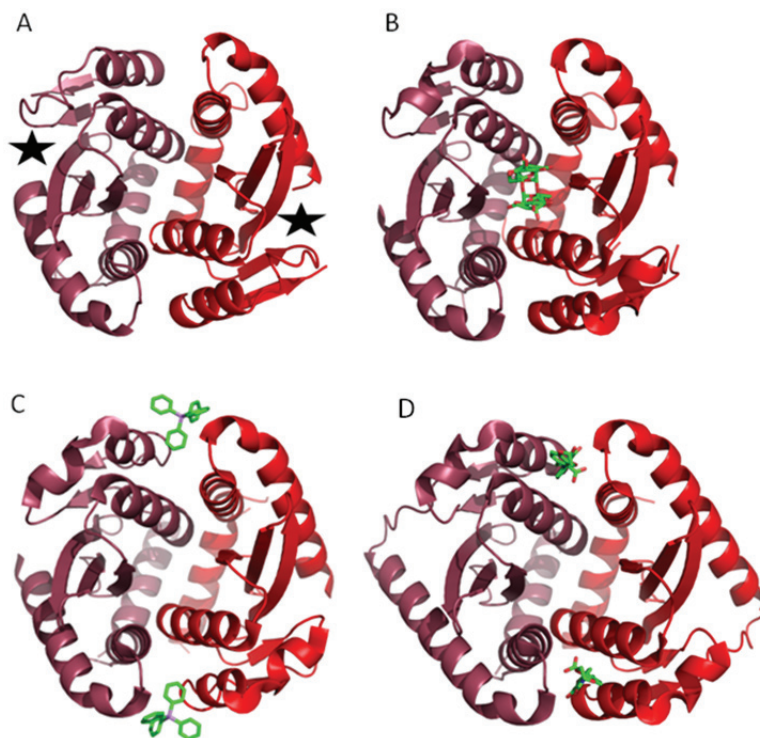


Figure 17 : Structures cristallines des CCD sous formes dimériques de l'intégrase du VIH-1 en complexe avec différents inhibiteurs allostériques potentiels. Les sites actifs sont localisés par une étoile noire sur la forme « apo » et les composés sont représentés en vert. (A) Forme « apo » (PDB 1BL3). (B) Forme complexée avec une molécule de sucrose (PDB 3L3V). (C) Forme complexée avec le tétraphényl arsonium (PDB 1HYV). (D) Forme complexée avec le composé « 20 » décrit par D. I. Rhodes et collaborateurs (PDB 3NFA).

Le concept opposé visant à forcer l'oligomérisation de l'intégrase en oligomères d'ordre élevé permettrait également d'inactiver l'enzyme. Pour cela des peptides dérivés de la protéine LEDGF ont été utilisés pour déplacer l'équilibre d'oligomérisation de l'intégrase du VIH-1 vers la formation de tétramères qui ne seraient pas capables de réaliser l'étape de clivage (Hayouka *et al.* 2007). Sur le même principe, les composés qui interagissent avec l'interface dimérique du CCD contribuent à stabiliser un état oligomérique particulier de l'enzyme (ici la forme dimérique) déplaçant ainsi l'équilibre de l'oligomérisation vers cette forme, au dépend de la forme tétramérique qui est nécessaire pour l'étape de transfert de brins (Demeulemeester *et al.* 2012).

De la même façon, des inhibiteurs non compétitifs de l'intégrase peuvent être développés et inhiber l'activité de cette enzyme. On peut citer l'exemple du composé Y-3 qui a été co-cristallisé avec le domaine central de l'intégrase du RSV (Lubkowski *et al.* 1998). Cette molécule aromatique ne se lie pas dans le site actif mais plutôt à sa surface. Bien que le Y3 ne se lie pas dans le site actif et ne semble pas influencer l'état d'oligomérisation du CCD, il influence la conformation d'une boucle flexible interférant directement avec la liaison de l'intégrase à l'ADN. Ce composé inhibe les activités de clivage et de transfert de brins de l'intégrase avec une bonne efficacité. On peut également citer l'exemple de l'utilisation de dérivés d'arsenic qui ont été co-cristallisés avec l'intégrase du VIH-1

(Molteni *et al.* 2001). Ces composés se lient à l'interface du dimère du cœur catalytique et inhibent les activités catalytiques de l'enzyme en agissant indirectement sur la conformation du site actif de l'enzyme (**figure 17C**). Cet effet a également été observé pour le composé « 20 » décrit par D.I. Rhodes et collaborateurs et semble interférer sur la conformation d'une boucle située à proximité du site actif du CCD d'IN VIH-1, inhibant ainsi ses activités (**figure 17D**) (Rhodes *et al.* 2011).

Une autre stratégie vise également au développement de composés inhibant l'interaction de l'intégrase avec ses partenaires viraux et cellulaires au sein du PIC (Demeulemeester *et al.* 2012). Altérer ces interactions devrait perturber le mécanisme d'intégration. Dans ce contexte, les inhibiteurs d'interaction entre l'intégrase du VIH-1 et son partenaire cellulaire LEDGF, ont fait l'objet de nombreuses études. Des composés peptidiques ou chimiques (appelés LEDGIN) interagissant spécifiquement avec l'intégrase au niveau de son site d'interaction avec LEDGF (en particulier au niveau du CCD) ont été identifiés et perturbent efficacement la réplication des rétrovirus (**figure 18**) (Christ *et al.* 2010; Rhodes *et al.* 2011; Kessl *et al.* 2012). La stratégie inverse a également été utilisée pour le VIH-1 et vise à utiliser des peptides de régions de l'intégrase connus comme interagissant avec LEDGF (Cavalluzzo *et al.* 2012).

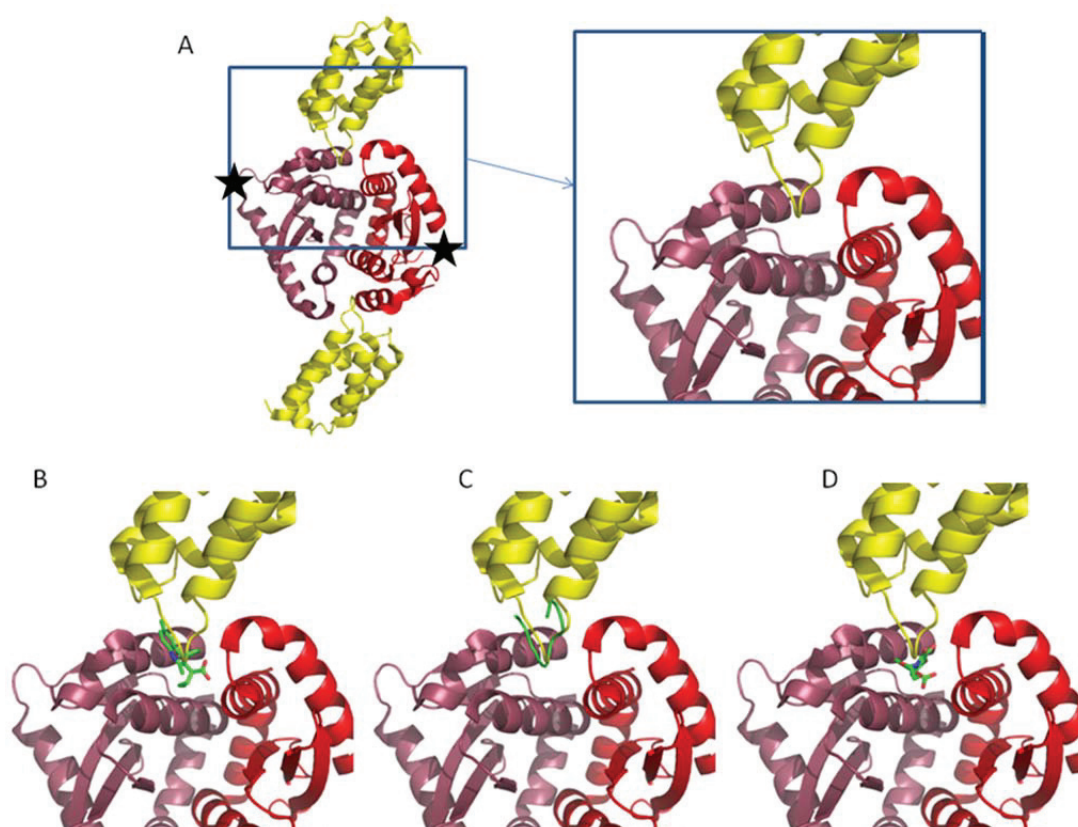


Figure 18 : (A) Structure cristalline « apo » du CCD sous forme dimérique de l'intégrase du VIH-1 en complexe avec IBD-LEDGF (CCD en rouge, sites actifs localisés par une étoile noire, IBD-LEDGF en jaune, PDB 2B4J). (B) Position du composé « LEDGIN 6 » (en vert) décrit par F. Christ et collaborateurs dans la structure 2B4J, après superposition de la structure complexée (PDB 3LPU) avec la structure « apo ». (C) Position du peptide « 4 » (en vert) décrit par D.I. Rhodes et collaborateurs dans la structure 2B4J, après superposition de la structure complexée (PDB 3AVA) avec la structure « apo ». (D) Position du composé « 20 » (en vert) décrit par D.I. Rhodes et collaborateurs dans la structure 2B4J, après superposition de la structure complexée (PDB 3NFA) avec la structure « apo ».

Le premier médicament ciblé contre l'intégrase, le Raltégravir, a démontré un effet antirétroviral rapide et efficace chez des patients infectés par le VIH-1. Il représente le premier

inhibiteur de l'intégrase a avoir été approuvé par la FDA (2007) pour son utilisation clinique (Demeulemeester *et al.* 2012). Toutefois, ce traitement s'accompagne rapidement par l'apparition de souches résistantes. D'autres inhibiteurs de l'intégrase et ciblant le site actif, l'Elvitégravir et le Dolutégravir, sont actuellement en fin de phases cliniques, mais semblent également déjà rencontrer des phénomènes de résistances (Mouscadet *et al.* 2010). Pour prévenir l'apparition de résistances croisées, de nouveaux inhibiteurs ciblés contre d'autres sites de l'intégrase (inhibiteurs allostériques) doivent continuer à être développés. La combinaison de plusieurs inhibiteurs de l'intégrase permettra d'aboutir à des conditions optimales de traitement pour lutter contre la prolifération du VIH-1 et l'apparition de multi-résistances, en complétant la multithérapie actuelle.

OBJECTIFS DE L'ÉTUDE

Mon travail de thèse a principalement porté sur l'étude des propriétés moléculaires et structurales de l'intégrase du rétrovirus de l'espèce aviaire RAV-1, qui constitue un modèle d'étude de l'intégrase du VIH-1. Cette dernière s'est déroulée au sein des équipes de « Biocristallographie et Biologie Structurale des Cibles Thérapeutiques » et « Rétrovirus, Intégration Virale et Vaccins » du pôle scientifique de Lyon-Gerland, sous la direction du Pr. P. Gouet et du Dr C. Ronfort.

Cette recherche s'est inscrite dans la continuité du travail de thèse effectué par le Dr A. Ballandras sur cette même enzyme de 2009 à 2012. Lorsque j'ai débuté mes recherches, Allison Ballandras venait de résoudre la structure cristalline du domaine catalytique (CCD) de l'intégrase du RAV-1 à une résolution de 1,8 Å. La structure dimérique du CCD obtenue révélait un « nouvel » assemblage, stabilisé à son interface intermoléculaire par trois paires d'hélices α contre deux paires d'hélices α dans tous les assemblages dimériques « canoniques » de CCD observés jusqu'à présent (RSV, HIV-1, SIV, PFV...). Le passage de l'interface dimérique « canonique » à la « nouvelle » interface nécessite un déplacement significatif de 15 Å d'un monomère le long du deuxième monomère. Les structures tertiaires des CCD sont préservées lors de ce réarrangement. L'analyse de ce « nouvel » assemblage montre que cette translation donne lieu à l'apparition de canaux symétriques à l'interface dimérique, délimités par les paires d'hélices $\alpha 1-\alpha 1$ et $\alpha 3-\alpha 5$. Une molécule d'acide 2-(N-Morpholino)-EthaneSulfonique (MES), utilisée comme tampon dans la solution de cristallisation, est fixée dans un de ces deux canaux et semble participer à la stabilisation de cette « nouvelle » interface. De même, l'interface est stabilisée en son centre par un ion zinc qui relie les groupements imidazoles de deux histidines symétriques (H103A et H103B) (**figure 19**).

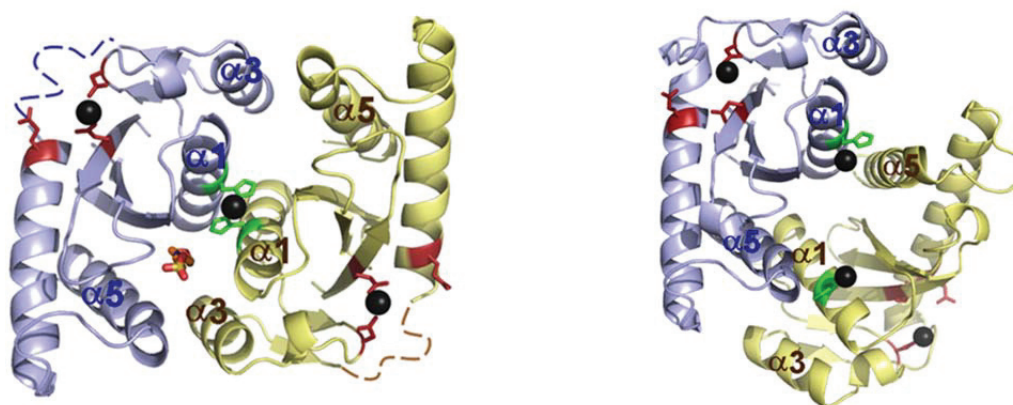


Figure 19 : (Gauche) Structure cristalline du CCD sous forme dimérique de l'intégrase du RAV-1 avec la nouvelle interface à trois paires d'hélices α (code PDB : 3O4N). Les histidines 103 sont représentées en vert, les résidus catalytiques de chaque site actif en rouge, la molécule de MES en orange et les ions zinc en noir.

(Droite) Structure cristalline du CCD sous forme dimérique de l'intégrase du RSV (forme canonique) avec l'interface à deux paires d'hélices α (code PDB : 1VSM).

L'article présentant ces résultats était en cours d'écriture lorsque j'ai commencé ma thèse. Une condition nécessaire à sa publication était de montrer que la « nouvelle » interface observée *in crystallo* pouvait se former en solution. En effet, des expériences réalisées par Allison Ballandras durant sa thèse ont montré que le domaine catalytique isolé de l'intégrase du RAV-1 était majoritairement sous forme monomérique en solution, même en présence de MES. De plus, l'ajout de zinc induisait la précipitation de la protéine ne permettant pas de mettre en évidence la « nouvelle » forme dimérique par les méthodes biochimiques classiques.

Une première partie de mon travail a donc porté sur l'étude du domaine catalytique isolé de l'intégrase du RAV-1 (IN-CCD). Une étude par modélisation moléculaire a permis de mettre en évidence la possibilité de former un pont disulfure inter-moléculaire à l'interface du « nouvel »

assemblage par la mutation de l'histidine 103 en cystéine. J'ai donc dans un premier temps dû surproduire le domaine catalytique muté de l'intégrase du RAV-1 (IN-CCD-H103C) en système procaryote (*E. coli*) et étudier son comportement en solution, avant de parvenir à l'assembler de manière stable selon la « nouvelle » interface. Ces recherches que j'ai menées et l'article qui en a découlé sur la structure du CCD de l'intégrase du RAV-1 (Ballandras *et al.* 2011) sont décrits en détail dans la suite du manuscrit.

Mon travail de thèse a ensuite consisté à rechercher des molécules susceptibles de stabiliser le domaine catalytique de l'intégrase de RAV-1 sous ce « nouvel » assemblage dans le but de tester par la suite l'effet de ces composés *in vitro* (activités catalytiques de l'intégrase) et *in cellulo* (sur le cycle réplcatif rétroviral). En effet, notre hypothèse de travail est que cette nouvelle association dimérique, qui semble significativement différente de celle observée pour la phase d'intégration, pourrait se produire sur l'enzyme entière et engendrer des réarrangements structuraux des autres domaines, la rendant inactive pour l'étape d'intégration. Ainsi, des composés analogues au MES et présentant une affinité élevée pour le nouvel assemblage dimérique, pourraient être utilisés en tant qu'inhibiteurs allostériques, pour perturber le cycle de réplication rétroviral. Une autre partie de mon travail a donc consisté à déterminer un protocole efficace de surproduction et de stabilisation de la nouvelle forme dimérique obtenue avec le mutant IN-CCD-H103C. Il m'a ensuite fallu mettre au point les techniques permettant de quantifier les interactions en solution entre le dimère stabilisé et des composés analogues au MES, le but étant d'identifier les meilleurs candidats par des expériences *in vitro* et d'obtenir des informations structurales sur ces complexes afin de développer rationnellement de nouveaux composés plus affins. En parallèle, j'ai également dû mettre au point des techniques permettant d'observer l'effet de ces composés sur la forme sauvage du domaine catalytique (IN-CCD-WT) et en particulier sur son état d'oligomérisation. Ces recherches que j'ai menées sur l'intégrase du RAV-1 ont été réalisées en collaboration avec le Dr Philippe Maurin de l'équipe de « Chimie Bio-organique » de l'Ecole Normale Supérieure (ENS) de Lyon, qui a synthétisé les molécules candidates.

Par ailleurs, nous avons observé que la formation de cette nouvelle interface dimérique permettait l'apparition d'un sillon fortement basique à sa surface (**figure 20**). Des expériences de *docking*, décrites dans l'article, montrent que ce sillon pourrait lier un acide nucléique simple brin. Le nouvel assemblage présenterait donc un rôle biologique potentiel si cette hypothèse était vérifiée. J'ai donc également mené durant ma thèse des recherches sur le rôle biologique potentiel de cette nouvelle interface et en particulier par des expériences de trempages de cristaux du domaine catalytique sauvage IN-CCD-WT avec divers nucléotides.

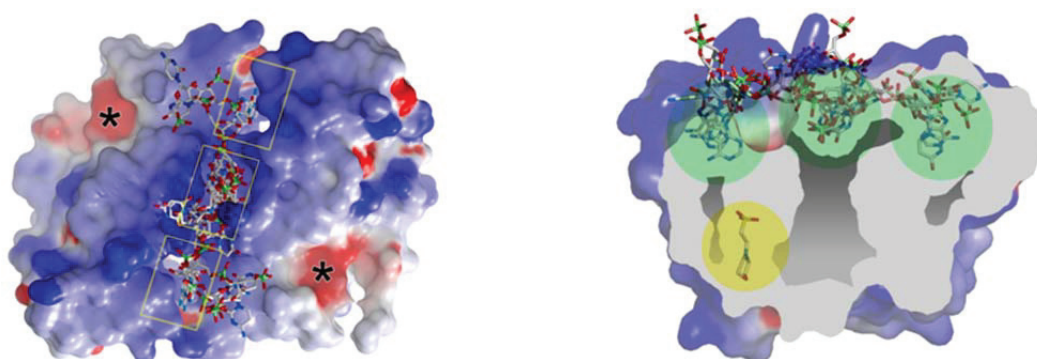


Figure 20 : (**Gauche**) Projection des potentiels électrostatiques sur la surface moléculaire de la nouvelle forme dimérique du CCD de l'intégrase du RAV-1 et modélisation d'un ARN simple brin fixé sur le sillon basique de l'interface.

(**Droite**) Plan de coupe à 90°. Le site actif est représenté par un astérisque noir. Le site de fixation du MES est visualisé en jaune. Le site de fixation des acides nucléiques prédit par *docking* est visualisé en vert.

Une seconde partie de mon travail a porté sur l'étude de l'intégrase entière du RAV-1 (IN-FL). Sa structure tridimensionnelle n'étant alors pas connue, j'ai aussi mis en place un protocole de surproduction et de purification d'IN-FL-WT afin de permettre d'initier des essais de cristallisation. J'ai également caractérisé son état d'oligomérisation en solution ainsi que son état de dispersité avant de mener une étude par diffusion des rayons X aux petits angles (SAXS), pour modéliser son enveloppe (collaboration avec le Dr Patrice Vachette de l'équipe « Conformation Des Macromolécules en Solution », IBBMC, Paris-Orsay). Le protocole mis au point pour la protéine entière sauvage a été utilisé pour surproduire et purifier la protéine entière mutée (IN-FL-H103C) dans le but de mettre en évidence l'existence de la « nouvelle » interface stabilisée par un pont disulfure au sein de la protéine entière. Ces travaux sont également décrits dans la suite du manuscrit.

MATÉRIEL ET MÉTHODES

1. Obtention et stockage des ADN

1.1. Souches bactériennes utilisées

Principe : Une souche modifiée d'*Escherichia coli* (souche DH5 α) est utilisée pour la production et le stockage de l'ensemble des ADN plasmidiques d'intérêt. Cette souche bactérienne est très utilisée pour les applications de clonage, et comporte des mutations sur certains gènes, permettant d'améliorer la stabilité des ADN recombinants (Taylor *et al.* 1993).

Protocole : Des bactéries DH5 α (*Invitrogen*) sont ici utilisées. Ces bactéries sont maintenues en culture sur des boîtes de Petri contenant un milieu de culture solide *Lysogeny Broth-agar* ou LB-agar (*Sigma*). Ce milieu est supplémenté avec un antibiotique approprié à la sélection des vecteurs plasmidiques. Les bactéries sont conservées à court terme sous cette forme près de trente jours à 4 °C.

Parallèlement, ces bactéries sont cultivées jusqu'à saturation dans un milieu de culture liquide LB (*Sigma*) contenant un antibiotique approprié à la sélection des vecteurs plasmidiques. Une fois supplémenté en glycérol (20 % V/V final, *Euromedex*), le milieu est directement congelé à -80 °C dans des cryotubes de 2 ml (*Greiner Bio-One*) contenant 1 ml de milieu. Les bactéries ainsi congelées (stock glycérol), sont conservées à cette température pendant une plus longue durée (plusieurs mois ou années).

1.2. Extraction des ADN plasmidiques par lyse alcaline

Principe : La méthode de la lyse alcaline est employée afin de récupérer les ADN plasmidiques d'intérêt à partir des bactéries (Birnboim & Doly 1979). Cette méthode est basée sur la dénaturation/renaturation différentielle des ADN plasmidiques et chromosomiques.

Protocole : L'extraction sélective de l'ensemble des ADN plasmidiques d'intérêt est réalisée par la méthode de la lyse alcaline à l'aide d'un kit commercial (GeneJET Plasmid Miniprep Kit, *Fermentas*). Des bactéries DH5 α contenant un plasmide d'intérêt sont cultivées dans un milieu de culture liquide LB supplémenté avec un antibiotique approprié à la sélection du vecteur. Selon le taux de répllication du plasmide, les bactéries sont mises en culture dans 1-5 ml (fort taux de répllication) ou 5-10 ml (faible taux de répllication) de milieu LB, dans un agitateur à 37 °C sur la nuit, pour atteindre la saturation du milieu de culture. Ce dernier est ensuite centrifugé à 6 800 g et à température ambiante pendant deux minutes afin de sédimenter les bactéries. Le surnageant est ensuite éliminé et les bactéries resuspendues dans 250 μ l d'une solution contenant un tampon Tris-(hydroxyméthyl)-aminométhane (Tris pH 8,0), de l'acide diamine-éthylène-tétraacétique (EDTA) ainsi que l'enzyme Ribonucléase A (RnaseA). L'EDTA va chélater les cations divalents contenus dans la membrane externe bactérienne permettant de la fragiliser. Il permet de plus de priver les nucléases bactériennes de leur cofacteur magnésium limitant la dégradation des ADN recherchés. L'enzyme RnaseA permet la dégradation des ARN bactériens libérés suite à l'étape de lyse. Une fois que les cellules sont homogénéisées dans la solution précédente, est ajouté 250 μ l d'une solution de lyse (contenant le détergent anionique dodécylsulfate de sodium ou SDS et de la soude), et le milieu est inversé quatre à six fois. Le détergent SDS permet la rupture de la membrane lipidique bactérienne

et la lyse des bactéries. Le pH très alcalin de la soude assure de façon concertée avec le SDS la dénaturation des macromolécules dont les protéines bactériennes, mais également l'ADN chromosomique et l'ADN plasmidique. Par la suite, 350 µl d'une solution de neutralisation (contenant de l'acétate de potassium) est ajouté et le milieu inversé quatre à six fois. L'acétate de potassium permet la neutralisation de la soude et le retour à la normalité du pH. Cette neutralisation aboutit à la précipitation et à la renaturation incorrecte des protéines ainsi que de l'ADN chromosomique qui, du fait de sa taille, ne peut pas se réhybrider correctement. L'acétate de potassium a également pour effet de faire précipiter le SDS associé aux lipides et aux protéines, ce qui aboutit à la formation d'un précipité contenant les complexes SDS-lipides-protéines-ADN chromosomique. L'ADN plasmidique, de plus petite taille, peut quant à lui se réhybrider correctement et rester en solution. La solution est clarifiée par centrifugation à 12 000 g et à température ambiante pendant cinq minutes. Le surnageant contenant l'ADN plasmidique (ainsi que de petits fragments d'ADN chromosomique et des traces d'ARN) est récupéré et le culot de débris bactériens éliminé.

Afin d'améliorer la qualité de l'échantillon d'ADN plasmidique, l'échantillon est ensuite déposé sur une colonne de silice et cette dernière centrifugée à 12 000 g et à température ambiante pendant une minute. La silice présente une forte affinité pour l'ADN (Vogelstein & Gillespie 1979). La colonne est ensuite rincée par deux lavages successifs avec 500 µl d'une solution d'éthanol éliminée par une centrifugation à 12 000 g et à température ambiante pendant une minute. Une centrifugation supplémentaire à 12 000 g et à température ambiante pendant une minute est ensuite réalisée à vide afin d'éliminer les traces résiduelles d'éthanol qui pourraient interférer sur des étapes ultérieures de traitement de l'ADN plasmidique. L'ADN plasmidique retenu sur la silice est ensuite élué par l'ajout de 25-50 µl d'eau stérile suivi d'une centrifugation de la colonne à 12 000 g et à température ambiante pendant deux minutes. L'ADN récupéré est alors dosé par spectroscopie d'absorption ultraviolet-visible puis stocké sous cette forme à -20 °C pour être conservé.

1.3. Purification des ADN

Principe : Dans certaines circonstances, il est nécessaire de réaliser une purification des ADN, comme par exemple suite à une réaction d'amplification en chaîne par polymérase (PCR) afin d'éliminer les traces de nucléotides, d'enzymes ou encore de sels dans la solution contenant l'ADN d'intérêt, ou encore pour récupérer sélectivement certains fragments d'ADN suite à une séparation réalisée par électrophorèse sur gel d'agarose. La purification des ADN est basée sur ses propriétés de fixation par la silice.

Protocole : La purification des ADN est réalisée à l'aide d'un kit commercial (*Nucleospin Gel and PCR Clean-up, Macherey-Nagel*). Le mélange de PCR (pour 100 µl) ou l'agarose excisé et contenant le fragment d'ADN d'intérêt (pour 100 mg) est mélangé avec 200 µl d'une solution contenant des sels chaotropiques qui vont permettre la dénaturation des protéines des échantillons sans pour autant altérer les propriétés de liaison de l'ADN à la silice. Dans le cas de l'excision sur gel, le mélange est chauffé à 50 °C pendant cinq à dix minutes afin de faire fondre l'agarose. Le mélange est ensuite déposé sur une colonne de silice, puis une étape de centrifugation est réalisée à 11 000 g et à température ambiante pendant trente secondes. Un lavage de la colonne avec 700 µl d'une solution d'éthanol est effectué afin d'éliminer les traces de sels et de protéines dénaturées. Puis, une étape de centrifugation est réalisée à vide à 11 000 g et à température ambiante pendant deux minutes afin d'éliminer les traces d'éthanol. L'ADN d'intérêt est ensuite élué de la colonne par l'ajout de 15 à 50 µl d'eau suivi d'une incubation d'une minute à température ambiante et d'une centrifugation à 11 000 g et à température ambiante pendant une minute. L'ADN récupéré est alors

dosé par spectroscopie d'absorption ultraviolet-visible puis stocké sous cette forme à -20 °C pour être conservé.

2. Manipulations enzymatiques des ADN

2.1. Ligature de fragments d'ADN

Principe : La ligature d'un fragment d'ADN ou insert dans un vecteur de clonage est assurée par une enzyme, l'ADN ligase du phage T4. Cette enzyme catalyse en présence d'ATP et de magnésium la formation de liaisons phosphodiester entre une extrémité 5' phosphate et une extrémité 3' hydroxylée (Weiss & Richardson 1967).

Protocole : Pour la ligature d'un fragment d'ADN dans un vecteur de clonage, un vecteur d'ADN linéarisé est mis en présence de l'insert (traité par les mêmes enzymes de restriction que pour le vecteur ADN) et de l'ADN ligase du phage T4 dans un tampon adapté et donné par le fournisseur (*New England Biolabs*). La réaction de ligature est réalisée dans un volume final de 10 à 15 µl contenant une unité de ligase et un rapport molaire vecteur:insert allant de 1:3 à 1:10. Le mélange réactionnel est tout d'abord incubé trente minutes à température ambiante puis sur la nuit à 14 °C avant d'être conservé à 4 °C en attendant des étapes ultérieures.

2.2. Digestion des ADN par les enzymes de restriction

Principe : La digestion des ADN est catalysée par des enzymes particulières appelées « enzymes de restriction », reconnaissant des séquences d'ADN spécifiques et qui sont vendues par différents fournisseurs (*Roche, New England Biolabs, Fermentas*).

Protocole : La réaction standard a lieu dans un volume final de 20 µl comprenant 2 µl de tampon de restriction (1 X final) dont la composition dépend selon les enzymes de restriction, 0,5 à 5 µg d'ADN cible et une unité d'endonucléase de restriction par µg d'ADN cible. La durée d'incubation est généralement d'une heure et est réalisée à 37 °C. Les enzymes sont ensuite inactivées par dénaturation thermique, par une incubation du milieu réactionnel à 65 °C pendant vingt minutes puis, ce dernier est mis de côté à 4 °C en attendant des étapes ultérieures.

2.3. Traitement des ADN par l'enzyme Dpn1

Principe : Il est parfois nécessaire d'éliminer les ADN parentaux présents dans les produits de réactions PCR. Pour cela, une enzyme particulière est utilisée, l'enzyme Dpn1 issue de l'organisme *Diplococcus pneumoniae* G41 (Geier & Modrich 1979). L'enzyme Dpn1 est une enzyme de restriction reconnaissant spécifiquement la séquence palindromique de quatre bases GATC et est capable de cliver cette séquence entre l'adénosine et la thymidine, uniquement dans le cas où l'adénosine est méthylée. Les souches d'*Escherichia coli* *dam*⁺ (dont les souches DH5α) codent une enzyme de type méthylase qui catalyse la méthylation de certaines séquences spécifiques d'ADN. Les ADN isolés à partir de ces souches seront donc méthylés et pourront être la cible de l'enzyme Dpn1. A l'inverse, les ADN amplifiés par la réaction de PCR sont amplifiés *in vitro* et ne sont donc pas méthylés. Ils ne sont donc pas la cible de l'enzyme DPN1. Cette enzyme permet donc de dégrader spécifiquement les ADN matriciels utilisés pour la PCR mais pas les ADN amplifiés.

Protocole : Pour le traitement des ADN par l'enzyme Dpn1, à 50 µl de milieu réactionnel de PCR, est ajouté une unité de l'enzyme Dpn1 (*Fermentas*). Le mélange est incubé pendant 1 h 30 à 37 °C puis le milieu est mis de côté à 4 °C en attendant des étapes ultérieures.

2.4. Amplification *in vitro* de l'ADN par PCR

Principe : La réaction d'amplification en chaîne par polymérase ou PCR (pour *Polymerase Chain Reaction*) est une technique permettant la synthèse spécifique et exponentielle *in vitro* d'une région d'ADN prédéterminée par l'utilisation d'une enzyme thermostable à activité ADN polymérase ADN dépendante (Mullis *et al.* 1986). Une réaction PCR comprend un ADN cible ou matriciel, une enzyme ADN polymérase ADN dépendante thermostable, deux amorces nucléotidiques constituant les deux extrémités de l'ADN à amplifier et s'hybridant respectivement aux brins parallèles et antiparallèles de cet ADN, des oligonucléotides, un tampon réactionnel adapté à l'enzyme et du magnésium. Le principe de l'amplification repose sur la répétition de trois processus: la dénaturation de l'ADN cible bicaténaire à une température élevée (environ 95 °C) afin d'obtenir des molécules d'ADN monocaténares, l'hybridation des amorces oligonucléotidiques complémentaires de l'ADN monocaténaire cible par une diminution de la température (généralement entre 50 et 70 °C) et l'élongation par une ADN polymérase ADN dépendante thermostable à partir des amorces à une température proche de 72 °C. Les produits de ce premier cycle sont ensuite dénaturés par la chaleur et les amorces sont à nouveau hybridées avec les brins d'ADN provenant du premier cycle d'amplification, chaque brin servant de matrice à l'ADN polymérase aboutissant à une amplification exponentielle de l'ADN.

Protocole : Pour les réactions de PCR, le milieu réactionnel standard a un volume de 50 µl et contient une unité de polymérase haute fidélité ou polymérase Phusion (*New England Biolabs*), 10 ng d'ADN matriciel, 125 ng de chaque amorce oligonucléotidique (*Sigma*), 10 µl de tampon 5 X nécessaire à l'activité de l'enzyme (*New England Biolabs*) et 1 µl de 10 mM dNTP (*New England Biolabs*). Les réactions de PCR sont réalisées dans un thermocycleur GeneAmp PCR System 9700 (*Applied Biosystems*) selon un programme contenant les quatre étapes suivantes : dénaturation de l'ADN matriciel à 95 °C pendant trente secondes, hybridation spécifique des amorces oligonucléotidiques pendant une minute à une température propre aux amorces utilisées (température de fusion ou T_m) et élongation à 72 °C par l'ADN polymérase. Le temps d'élongation est fonction de la taille du fragment à amplifier et de la vitesse de polymérisation de la polymérase utilisée. Ces trois étapes sont répétées un nombre défini de cycles de l'ordre d'une dizaine. Enfin, suite au dernier cycle, une élongation finale est réalisée à 72 °C pendant quinze minutes. Suite à la réaction, le milieu réactionnel est stocké à 4 °C en attendant des étapes ultérieures.

2.5. Mutagenèse dirigée par PCR

Principe : La mutagenèse dirigée par PCR est une méthode permettant de générer de façon ciblée différents types de mutations (substitutions, insertions ou délétions) dans une séquence d'ADN plasmidique bicaténaire sans nécessité de sous clonages. Cette méthode repose sur l'observation qu'une stricte complémentarité de séquence de l'amorce nucléotidique avec l'ADN cible, n'est pas absolument nécessaire sur toute la longueur considérée pour la réaction de PCR. La procédure implique une réaction de PCR utilisant un ADN plasmidique bicaténaire comme ADN matriciel et deux amorces oligonucléotidiques chevauchantes contenant la mutation désirée (Papworth *et al.* 1996). Suite à la PCR, l'ADN matriciel, qui est méthylé, est digéré par l'enzyme Dpn1.

Le plasmide muté, qui est le produit de la PCR, n'est pas digéré et reste intact dans la solution (**figure 21**).

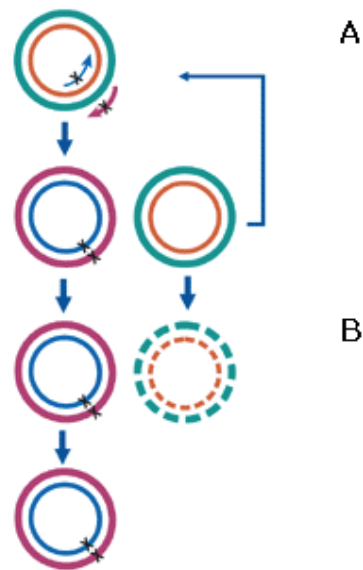


Figure 21 : Méthode de mutagenèse dirigée par PCR (Papworth *et al.* 1996). La synthèse des brins d'ADN comportant la mutation (**A**) est suivie de la digestion de l'ADN matriciel par l'enzyme DPN1 ne laissant intact que l'ADN amplifié et comportant la mutation (**B**) d'après (Smith 2007).

Protocole : Pour la mutagenèse dirigée par PCR, le protocole suivi est le même que le protocole classique de PCR avec une variation dans le nombre de cycles selon le type de mutation introduite (douze cycles pour la mutation d'une seule base, seize cycles pour une substitution et dix-huit cycles pour les délétions et insertions) (Papworth *et al.* 1996). La présence de la ou des mutations est ensuite vérifiée par séquençage.

3. Introduction des ADN dans les bactéries

Principe : La technique du choc thermique est utilisée pour transformer des bactéries *E. coli* préalablement rendues compétentes, par les ADN plasmidiques d'intérêt. Les bactéries utilisées dans ce travail sont déjà préalablement rendues compétentes par traitement chimique. La méthode du choc thermique consiste à mettre l'ADN plasmidique d'intérêt en présence des bactéries compétentes et à soumettre ces bactéries à un rapide changement de température. Les bactéries ainsi fragilisées incorporent les ADN plasmidiques et peuvent être sélectionnées en culture par l'utilisation d'un antibiotique approprié.

Protocole : Pour réaliser l'introduction des ADN plasmidiques dans les bactéries, de façon générale, à 50 µl de bactéries compétentes est ajouté 10 à 100 ng d'ADN plasmidique. Les cellules sont incubées trente minutes dans la glace puis trente secondes à 42 °C. Les bactéries sont ensuite placées deux minutes dans la glace afin de réaliser le choc thermique. Puis, les cellules sont reprises dans 250 µl de milieu de culture liquide LB et le mélange incubé à 37 °C pendant une heure sans antibiotique. Pour la sélection des bactéries transformées, 50 à 200 µl du milieu de transformation sont étalés sur des boîtes de Petri contenant un milieu de culture solide LB-agar et supplémenté avec

les antibiotiques appropriés. Les boîtes de Petri sont disposées à 37 °C sur la nuit avant d'être stockées à 4 °C.

4. Antibiotiques utilisés

Principe : Différents antibiotiques sont utilisés afin d'assurer la sélection des bactéries porteuses des ADN plasmidiques d'intérêt. Les antibiotiques utilisés dans ce travail sont le chloramphénicol, l'ampicilline et la kanamycine.

Le chloramphénicol est un antibiotique à large spectre qui appartient à la famille des phénicolés. Il bloque la synthèse des protéines en se fixant à la grande sous unité des ribosomes des procaryotes inhibant ainsi l'activité peptidyl transférase. La résistance au chloramphénicol est le plus souvent conférée par un gène codant une enzyme non sécrétée à activité acétyltransférase. Cette enzyme va catalyser l'ajout d'un groupement acétyl provenant de l'acétyl-coenzyme A sur le groupement hydroxyle du chloramphénicol permettant sa détoxification.

L'ampicilline est un antibiotique à large spectre qui appartient à la famille des bêta-lactamines. Cet antibiotique agit comme inhibiteur compétitif d'une enzyme à activité transpeptidase, impliquée dans la formation de la paroi cellulaire des bactéries gram positives. La résistance à l'ampicilline est le plus souvent apportée par un gène codant une enzyme à activité β -lactamase et sécrétée dans le milieu. Cette enzyme hydrolyse le cycle β -lactame de l'ampicilline désactivant ainsi ses propriétés antibiotiques.

La kanamycine est un antibiotique à large spectre qui appartient à la famille des aminosides. Cet antibiotique se fixe à la petite sous unité des ribosomes de procaryotes interférant dans la traduction des ARN messagers et entraînant l'arrêt de la traduction et la synthèse de protéines tronquées. La résistance à la kanamycine est le plus souvent apportée par un gène codant une enzyme non sécrétée à activité aminophosphotransférase. Cette enzyme catalyse le transfert d'un phosphate γ d'un ATP sur un hydroxyle de la kanamycine permettant sa détoxification.

Protocole : L'antibiotique chloramphénicol (*Sigma*) est préparé à 25 mg/ml dans une solution de 100 % V/V éthanol et est utilisé à une concentration finale de 25-50 μ g/ml dans les cultures de bactéries portant le gène de résistance au chloramphénicol. L'antibiotique ampicilline (*Sigma*) est préparé à 50-200 mg/ml dans de l'eau pour être utilisé à une concentration finale de 50-200 μ g/ml dans les cultures de bactéries portant le gène de résistance à l'ampicilline. L'antibiotique kanamycine (*Sigma*) est préparé à 50 mg/ml dans de l'eau pour être utilisé à une concentration finale de 25-50 μ g/ml dans les cultures de bactéries portant le gène de résistance à la kanamycine.

5. Surproduction des protéines recombinantes

5.1. Souches bactériennes utilisées

Principe : Différents variants d'une souche modifiée d'*Escherichia coli* (souche BL21), sont utilisés pour la production des protéines recombinantes d'intérêt.

La souche BL21 DE3 est couramment utilisée pour les expressions de protéines recombinantes et présente l'avantage d'être déficiente pour les protéases lon3 et ompT qui peuvent être responsables de la dégradation des protéines surproduites. Le terme « DE3 » indique que ces

bactéries ont subi une modification de leur génome dans lequel a été inséré le lysogène *DE3* placé sous le contrôle du promoteur de l'opéron lactose modifié pour le rendre peu sensible au phénomène de répression catabolique par le glucose (promoteur *lacUV5*) (**figure 22**). Le lysogène *DE3* code une ARN polymérase ADN dépendante provenant du bactériophage T7 ou PolT7. En présence de lactose ou d'un analogue non hydrolysable du lactose comme l'isopropyl β -D-1-thiogalactopyranoside ou IPTG, il y a levée de l'inhibition du répresseur LacI sur le site opérateur de l'opéron lactose permettant la transcription du gène de la polymérase T7 par l'ARN polymérase ADN dépendante bactérienne.

Pour permettre la surproduction d'une protéine d'intérêt, une stratégie développée par W. F. Studier consiste à transformer ces bactéries avec un gène d'intérêt placé sous le contrôle du promoteur reconnu par la PolT7, qui présente l'avantage d'être un promoteur fort, ainsi que de l'opérateur de l'opéron lactose (Davanloo *et al.* 1984; Studier & Moffatt 1986). Le gène d'intérêt remplace ainsi les gènes *LacZ*, *LacY* et *LacA* de l'opéron lactose. L'ajout de lactose ou d'un analogue non hydrolysable permet de lever la répression de LacI sur les opérateurs de l'opéron lactose et d'initier l'expression du lysogène *DE3* par l'ARN polymérase ADN dépendante bactérienne. L'enzyme PolT7 issue du lysogène *DE3* peut ainsi assurer la transcription du gène d'intérêt placé sous le contrôle du promoteur spécifique de la PolT7 et de l'opérateur de l'opéron lactose (**figure 23**).

Dans certains cas comme lorsque les protéines surexprimées sont toxiques pour les bactéries, il est nécessaire d'ajouter un élément de contrôle supplémentaire permettant de limiter les fuites d'expression du gène d'intérêt avant l'ajout de lactose ou d'analogues non hydrolysables du lactose. Les souches BL21 *DE3* pLysS contiennent un plasmide pLysS (chloramphénicol résistant) dans lequel un gène exprime le lysozyme du bactériophage T7 de façon constitutive. Le lysozyme T7 est un inhibiteur naturel de la polymérase T7 et son expression basale dans ces cellules permet de réduire les fuites d'expression des gènes d'intérêt liées à l'expression basale du gène de la PolT7 (figure 21) (Studier 1991). La production basale du lysozyme T7 a peu de conséquences sur la surexpression des gènes d'intérêt en présence d'inducteur. Une autre stratégie pour la surproduction de protéines toxiques consiste à utiliser des souches particulières dénommées souches « C41 *DE3* pLysS » et « C43 *DE3* pLysS ». Ces souches possèdent des mutations non caractérisées qui facilitent la surproduction de protéines toxiques et/ou membranaires chez *E. coli* (Miroux & Walker 1996).

L'expression de gènes hétérologues chez *E. coli* est parfois difficile, quand l'utilisation des codons dans le gène recombinant est différente de celle des cellules hôtes. La traduction de l'ARN recombinant est alors ralentie, ce qui aboutit à des protéines tronquées ou plus rarement à des modifications dans la séquence d'acides aminés de la protéine surproduite. Le problème est particulièrement connu pour les codons CGA (arginine), CGG (arginine), AGA (arginine), AGG (arginine), AUA (isoleucine), CUA (leucine), CCC (proline) et GGA (glycine). Ce problème peut être résolu par l'utilisation d'une souche particulière appelée « Rosetta 2 *DE3* pLysS » (Novy *et al.* 2001). Celle-ci dérive de la souche BL21 *DE3* pLysS mais possède au lieu du plasmide pLysS un plasmide pLysSRARE2 (chloramphénicol résistant) codant les ARNt des codons rares précédemment cités (sauf CGA) ainsi que le lysozyme T7.

La souche Rosetta-gami 2 *DE3* pLysS combine les avantages des souches Rosetta 2 *DE3* pLysS avec la possibilité de favoriser la formation des ponts disulfures à l'intérieur du cytoplasme des bactéries. Le milieu intra-cellulaire d'*E. coli* est normalement réducteur et certaines protéines recombinantes nécessitent, pour être correctement repliées, la formation de ponts disulfures intra ou inter-moléculaires. Cette souche comprend des mutations dans les gènes *trxB* et *gor* dont les produits sont responsables du contrôle du potentiel oxydo-réducteur du cytoplasme d'*E. coli* et la mutation de ces gènes favorise la formation des ponts disulfures intra-cellulaires (Bessette *et al.* 1999).

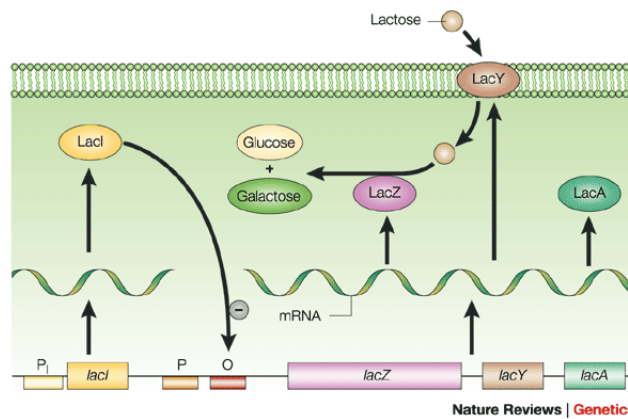


Figure 22 : Fonctionnement de l'opéron lactose chez *E. coli*. (Shuman & Silhavy 2003).

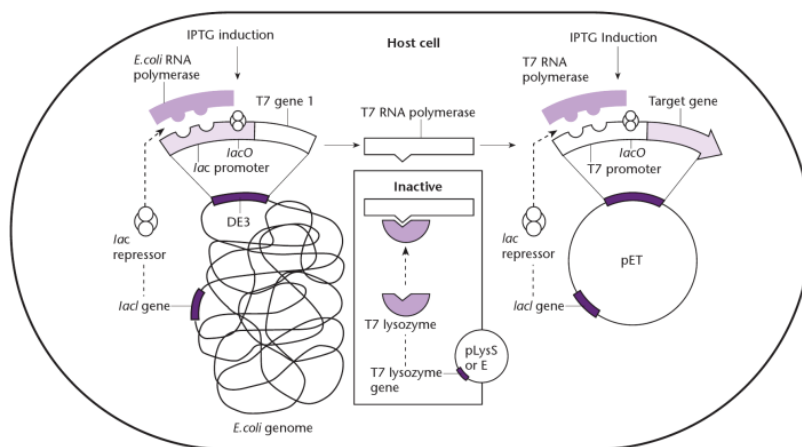


Figure 23 : Stratégie de surexpression de protéines recombinantes chez *E. coli* par le système de W. F. Studier (Studier & Moffatt 1986; Studier 1991) d'après (Primrose & Twyman 2006).

Les cellules BL21 sont également compatibles avec un autre système d'induction de la surproduction de protéines recombinantes, basé sur l'utilisation de promoteurs hybrides nommés promoteurs « *trc* » (constitué de la région -35 du promoteur tryptophane *trp* et de la région -10 du promoteur lactose modifié *lacUV5* séparés par dix-sept paires de bases) et promoteurs « *tac* » (constitué de la région -35 du promoteur tryptophane *trp* et de la région -10 du promoteur lactose modifié *lacUV5* séparés par seize paires de bases) (Amann *et al.* 1983).

Les promoteurs *trc* et *tac* sont reconnus par l'ARN polymérase ADN dépendante bactérienne. Ils sont plus faibles que le promoteur T7 mais plus fort que le promoteur *lac* sauvage. L'utilisation de ce type de promoteur peut être intéressante lorsque la production trop importante d'une protéine recombinante s'accompagne de problèmes de solubilité ou de toxicité.

Protocole : Différentes bactéries d'expression (BL21 DE3, BL21 DE3 pLysS, Rosetta 2 DE3 pLysS, Rosetta-gami 2 DE3 pLysS, C41 DE3 pLysS et C43 DE3 pLysS, *Euromedex*) ont été testées pour la surproduction des protéines recombinantes. Ces bactéries transformées par les ADN plasmidiques d'intérêt sont maintenues en culture sur des boîtes de Petri contenant un milieu de culture solide LB-agar (*Sigma*) supplémenté avec un antibiotique approprié à la sélection du vecteur plasmidique. Les bactéries sont conservées à court terme sous cette forme près de trente jours à 4 °C. Parallèlement, ces bactéries sont mises en culture jusqu'à saturation dans un milieu liquide LB (*Sigma*) contenant un

antibiotique approprié. Une fois supplémenté en glycérol à hauteur de 20 % V/V final (*Euromedex*), le milieu est directement congelé à -80 °C dans des cryotubes de 2 ml (*Greiner Bio-One*) contenant 1 ml de milieu. Les bactéries ainsi congelées (on parle de stock glycérol) sont conservées à cette température pendant une plus longue durée (plusieurs mois ou années).

5.2. Vecteurs plasmidiques utilisés

Différents vecteurs plasmidiques ont été utilisés pour la surproduction des protéines recombinantes chez *E. coli* (**figure 24**).

Plasmide pET-30a : Le plasmide pET-30a (*Novagen*) possède l'origine de réplication M13 qui permet la production de particules virales contenant l'ADN plasmidique sous la forme d'un ADN monocaténaire lors d'une infection par le bactériophage M13. Il possède également l'origine de réplication bactérienne ColE1 nécessaire à la réplication de l'ADN plasmidique dans les bactéries. Le gène *KanR* confère aux bactéries transformées par ce plasmide une résistance à l'antibiotique kanamycine. Le gène codant le répresseur LacI, le promoteur T7, l'opérateur de l'opéron lactose ainsi que le terminateur T7 permettent la surproduction de protéines recombinantes basée sur le système de W. F. Studier. Les protéines recombinantes peuvent selon l'insertion de leur gène dans le site multiple de clonage porter une étiquette amino-terminale de configuration six histidines/site de clivage par la thrombine/étiquette S/site de clivage par l'entérokinase ou une étiquette carboxy-terminale six histidines. L'étiquette six histidines permet de faciliter la purification de la protéine recombinante par un système de chromatographie d'affinité pour le métal (Porath *et al.* 1975). L'étiquette S qui est un oligopeptide de quinze acides aminés issue de la ribonucléase A pancréatique peut améliorer la solubilité de la protéine surproduite (Raines *et al.* 2000). Les étiquettes du côté amino-terminal pourront être clivées par l'utilisation des enzymes entérokinase ou thrombine (Chang 1985; Choi *et al.* 2001).

Ce plasmide a été utilisé pour la surproduction de l'intégrase entière du virus RAV-1 et le gène recombinant a été cloné entre les sites NcoI et EcoRI permettant la surproduction d'une protéine étiquetée du côté amino-terminal.

Plasmide pPROEX-HTa : Le plasmide pPROEX-HTa (*Life Technologies*) possède les origines de réplication M13 ainsi que ColE1 et la présence du gène *amp* confère aux bactéries transformées par ce plasmide, une résistance à l'antibiotique ampicilline. Le gène codant le répresseur LacI, le promoteur *trc*, l'opérateur de l'opéron lactose ainsi que les terminateurs de transcriptions bactériens *rrnB* et *rrnB* T2, permettent la surproduction de protéines recombinantes basée sur le système du promoteur bactérien hybride *trc*. L'insertion d'un gène recombinant dans le site multiple de clonage aboutit à la production d'une protéine possédant une étiquette du côté amino-terminal avec une configuration six histidines/site de clivage par la protéase du TEV (Parks *et al.* 1994).

Ce plasmide a été utilisé pour la surproduction de l'intégrase entière du virus RAV-1 (sauvage et H103C) et le gène recombinant a été cloné entre les sites NcoI et EcoRI permettant la surproduction d'une protéine étiquetée du côté amino-terminal.

Plasmide pETG-10A : Le plasmide pETG-10A (*EMBL*) possède les origines de réplication M13 ainsi que ColE1 et la présence du gène *amp* confère aux bactéries transformées par ce plasmide une résistance à l'antibiotique ampicilline. Le gène codant le répresseur LacI, le promoteur T7, l'opérateur de l'opéron lactose ainsi que le terminateur T7 permettent la surproduction de protéines recombinantes basée sur le système de W. F. Studier. La particularité de ce plasmide réside dans l'utilisation du système Gateway (*Invitrogen*) basé sur l'insertion d'un gène recombinant entre les sites AttR2 et AttR1 par un système de recombinaison homologue utilisé chez le phage λ (Hartley *et al.* 2000). La recombinaison du gène d'intérêt entre ces sites va permettre de perdre la résistance au chloramphénicol apportée par le gène *CmR* ainsi que le gène *ccdb* exprimant une protéine toxique pour *E.coli*. L'insertion d'un gène recombinant entre AttR2 et AttR1 aboutit à la production d'une protéine possédant une étiquette du côté amino-terminal avec une configuration 6 histidines.

Ce plasmide a été utilisé pour la surproduction du domaine catalytique de l'intégrase du virus RAV-1 (sauvage et H103C) dont le gène a été modifié afin de comporter un site de clivage par la thrombine avant l'étiquette six histidines présente du côté amino-terminal.

Plasmide pRK793 : Le plasmide pRK793 (Kapust *et al.* 2001) possède les origines de réplication M13 ainsi que ColE1 et la présence du gène *amp* confère aux bactéries transformées par ce plasmide une résistance à l'antibiotique ampicilline. Le gène codant le répresseur LacI, le promoteur *tac*, l'opérateur de l'opéron lactose ainsi que les terminateurs de transcriptions bactériens *rrnB* et *rrnB* T2 permettent la surproduction de la protéase du TEV (mutant S219V) basée sur le système du promoteur bactérien hybride *tac*. La protéase surproduite possède une étiquette amino-terminale de configuration MBP/site de clivage du facteur X/site de clivage par la protéase du TEV/étiquette six histidines et une étiquette carboxy-terminale contenant plusieurs arginines. Les étiquettes MBP et poly-arginine permettent d'améliorer la solubilité de la protéase surproduite (Sassenfeld & Brewer 1984; Di Guan *et al.* 1988). L'existence d'un site de clivage par la protéase du TEV entre la protéine MBP et la protéase TEV permet son autoclivage *in vivo* au cours de la surproduction. Le mutant S219V empêche l'autoinactivation de la protéase (Kapust *et al.* 2001).

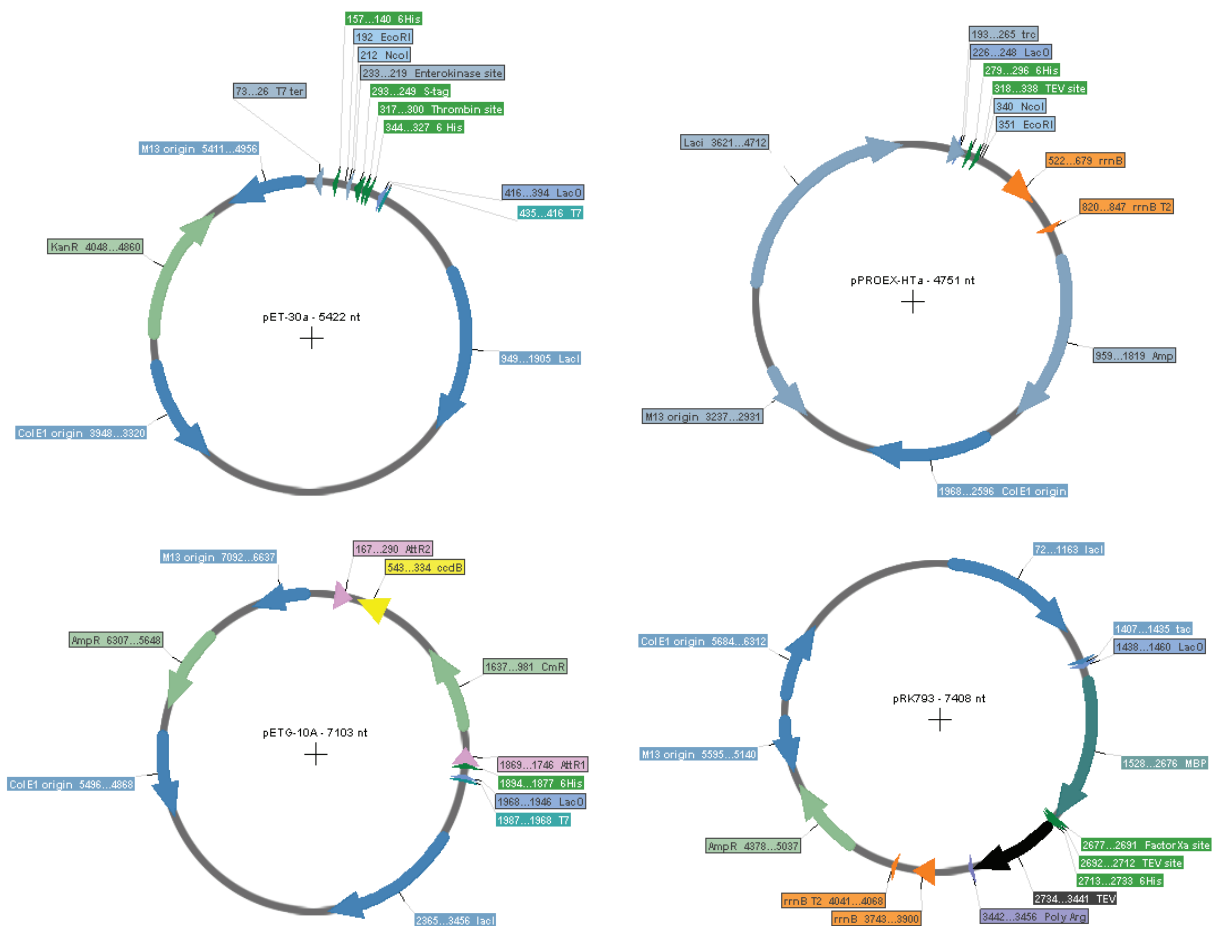


Figure 24 : Représentations simplifiées des vecteurs plasmidiques pET-30a (**haut, gauche**), pPROEX-HTa (**haut, droite**), pETG-10a (**bas, gauche**) et pRK793 (**bas, droite**).

5.3. Méthodes de culture et surproduction des protéines recombinantes

Principe : La surproduction des protéines est réalisée selon différents protocoles suivant les protéines étudiées. Les étapes générales de culture des bactéries et d'induction de la surproduction des protéines recombinantes suivent des règles similaires.

Protocole : Des bactéries adaptées à la surproduction de protéines recombinantes sont transformées par les plasmides d'intérêt. Les bactéries sont ensuite mises en culture dans un agitateur à 37 °C sur la nuit dans un milieu de culture liquide LB supplémenté en antibiotiques, directement à partir des stocks glycérol ou à partir d'un clone isolé sur un milieu de culture solide LB-agar contenant les antibiotiques appropriés. Une fois à saturation, cette pré-culture est utilisée pour inoculer un plus grand volume de milieu contenant les antibiotiques appropriés. Facultativement, les clones d'un milieu de culture solide peuvent être repris dans un milieu de culture liquide, qui sera utilisé comme pré-culture afin d'inoculer un plus grand volume (*Plating method*) (Suter-Crazzolaro & Unsicker 1995).

La culture est ensuite incubée à 37 °C sous agitation (180 à 220 rpm) et la croissance bactérienne est suivie par mesure de la turbidité du milieu à l'aide d'un spectrophotomètre jusqu'à atteindre une absorbance désirée (absorbance entre 0,5 et 1 pour une longueur d'onde de 600 nm) qui correspond généralement à la phase exponentielle de la croissance des bactéries pour un milieu

LB. La surproduction des protéines d'intérêt est alors induite par l'ajout d'IPTG (*Euromedex*) dans le milieu de culture pendant une durée variable. Les cultures sont ensuite centrifugées à 6 000 g et à 4 °C pendant dix minutes afin de sédimenter les bactéries. Le surnageant est ensuite éliminé puis les culots bactériens stockés à -20 °C en attendant l'étape de lyse des bactéries.

Pour certaines protéines étudiées, différentes concentrations d'IPTG (0,3-1 mM final) ainsi que différents temps (quatre heures ou sur la nuit) et températures d'induction (25 ou 37 °C) sont testés afin d'améliorer la quantité et la solubilité des protéines surproduites. Pour certains essais, une solution d'Augmedium (1 X final, *AthenaES*) est ajoutée dans la culture et avant l'induction afin d'activer la surproduction des chaperonnes endogènes de *E. coli*. Deux autres milieux de culture (Turbo et Hyper, *AthenaES*) ont également pu être testés comme alternatives au milieu LB.

6. Purification des protéines recombinantes

6.1. Extraction des protéines

Principe : L'extraction des protéines bactériennes nécessite une lyse efficace des cellules mais également l'utilisation d'une solution tampon adaptée à la protéine d'intérêt pour permettre son extraction efficace de la fraction insoluble (débris bactériens). Les conditions optimales d'extraction (méthode de lyse et nature des solutions tampons) vont pouvoir varier selon les protéines étudiées et doivent parfois être optimisées. Il existe différentes méthodes de lyse possibles (mécaniques ou chimiques/enzymatiques) et qui sont plus ou moins adaptées selon les protéines étudiées. Les méthodes de lyse qui sont employées dans ce travail sont la lyse par sonication et la lyse par le système microfluidique (méthodes mécaniques). L'extraction des protéines s'effectue dans une solution tampon (ou tampon de lyse) de composition variable selon les protéines d'intérêt et dans laquelle les cellules bactériennes ont été préalablement resuspendues.

La lyse par sonication consiste en la destruction des cellules par l'utilisation d'ondes sonores libérées par une sonde ultrasonique immergée dans la solution de resuspension des bactéries. Ces ondes ultrasonores causent la croissance de bulles microscopiques par un phénomène dit de cavitation gazeuse. Ces bulles vont imploser sous l'effet des ondes avec pour conséquences une augmentation locale brutale de pression et de température dans le champ de cavitation aboutissant à la destruction des cellules.

La lyse par le système microfluidique est basée sur l'utilisation d'écoulements à haute pression qui entrent en collision à des vitesses extrêmement élevées à l'intérieur de micro-canaux définis avec précision. Les forces combinées de cisaillement et d'impact des particules vont aboutir à la destruction des cellules.

Protocole : Pour la méthode de sonication, les cellules (pour un litre de culture) sont resuspendues dans une solution tampon (20 à 40 ml) adaptée à la protéine étudiée, et soniquées à l'aide d'un appareil Sonifier 450 (*Branson*) par l'application de cycles de trente secondes de sonication/trente secondes de refroidissement dans la glace pendant une durée totale de sonication de près de trois à six minutes (réglages *duty cycle* 40 à 50 %, *output control* 4 à 5).

Pour la méthode de la lyse par système microfluidique, les cellules (pour un litre de culture) sont resuspendues dans une solution tampon (20 à 40 ml) adaptée à la protéine étudiée. La solution

de bactéries est injectée dans un appareil Microfluidizer M-110P (*Microfluidics*) dont le circuit est refroidi dans la glace (quatre à six passages successifs à 18 000 psi de pression suffisent).

La lyse des cellules s'accompagne de la libération d'ADN bactérien mais également de protéases cytosoliques et membranaires qui pourraient dégrader les protéines d'intérêt extraites. Les ajouts de Dnase1 ainsi que d'inhibiteurs d'antiprotéases dans la solution tampon s'avèrent donc nécessaire. Les solutions tampons sont donc supplémentées en Dnase1 (20 µg/ml, *Sigma*) et MgCl₂ (nécessaire à l'activité de la Dnase 1, 10 mM, *Sigma*) mais également d'un cocktail d'antiprotéases sans EDTA (*Complete protease inhibitor cocktail EDTA free, Roche*).

Les lysats cellulaires sont ensuite clarifiés par centrifugation à 14 000 g et à 4 °C pendant trente minutes à une heure afin de séparer la fraction soluble (surnageant) et la fraction insoluble (débris bactériens). Le surnageant est ensuite filtré sur un filtre avec une porosité de 0,45 µm (*WVR International*) avant de passer à l'étape de purification.

6.2. Purification par chromatographie

6.2.1. Chromatographie d'affinité pour le métal

Principe : L'ensemble des protéines étudiées dans ce travail est purifié par le système de chromatographie d'affinité pour le métal (*Immobilized Metal ion Affinity Chromatography* ou IMAC) et plus spécifiquement pour le nickel développé par J. Porath et collaborateurs (Porath *et al.* 1975). La purification des protéines d'intérêt par ce système est possible par la présence d'une étiquette poly-histidines à l'une des extrémités de la protéine recombinante. Cette étiquette contenant le plus souvent six histidines successives, peut établir des liaisons de coordination avec un cation métallique divalent (généralement de nickel) immobilisé sur un support solide par l'intermédiaire d'un bras espaceur, comme l'acide iminodiacétique (IDA) ou l'acide nitrilotriacétique (NTA). L'ion nickel possède six liens de coordination. Trois sont sollicités par le bras espaceur IDA contre quatre pour le bras espaceur NTA permettant d'immobiliser l'ion nickel sur la résine. Il en reste trois pour le bras IDA et deux pour le bras NTA qui peuvent interagir avec les atomes d'azote du cycle de la chaîne latérale de deux histidines (**figure 25**). L'élution de la protéine se fait à l'aide d'un composé qui va entrer en compétition avec l'histidine : l'imidazole libre ou par l'ajout d'EDTA qui va chélater les ions nickel de la résine.

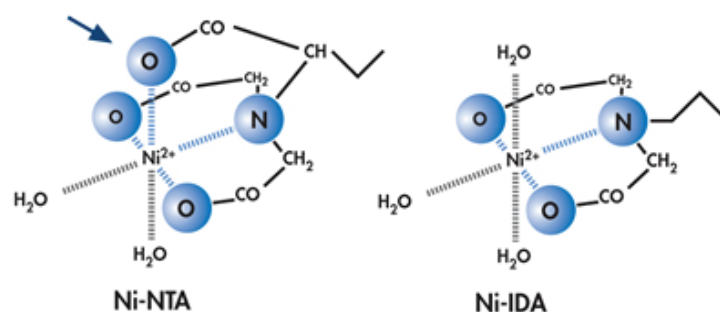


Figure 25 : Structures des bras NTA (**gauche**) et IDA (**droite**) complexés avec le nickel et utilisés dans le système IMAC d'après www.qiagen.com.

Protocole : La purification des protéines par le système IMAC est réalisée à l'aide d'un système de chromatographie liquide des protéines à basse pression (*Fast Protein Liquid Chromatography* ou FPLC, système *Aktä purifier*, *GE Healthcare*). Une colonne est reliée à un système permettant aux solutions tampons de la traverser de haut en bas. Ce dernier comprend deux pompes, P1 et P2, reliées aux solutions de tampon T1 ou T2 (pour P1) et T3 (pour P2). Ces pompes vont amener les tampons dans une chambre de mélange rendant possible la création d'un gradient. Le mélangeur est connecté au distributeur qui va assurer l'arrivée de tampon dans la colonne. L'échantillon à chromatographier est également déposé dans le distributeur. En aval de la colonne est installé un conductimètre (qui enregistre la conductance du tampon dépendant de sa composition) et une cellule spectrophotométrique qui enregistre l'absorbance aux ultraviolets (généralement à 280 nm). L'éluat de la colonne est ensuite recueilli par petites fractions par un collecteur. Pour réaliser les purifications par le système IMAC, les solutions de surnageants filtrées récupérées après la lyse des bactéries et contenant les protéines d'intérêt sont injectées dans une boucle d'injection reliée au distributeur. Des colonnes de 1 ml chélatées avec du nickel par l'intermédiaire d'un bras espaceur de type IDA sont utilisées (Protino Ni-IDA 1 ml, *Macherey Nagel*). Après le passage de la totalité du surnageant sur la colonne pré-équilibrée avec le tampon de lyse, un premier lavage est réalisé avec le tampon de lyse ou tampon T1 afin d'éluer les protéines qui ne se sont pas fixées sur la colonne. Un deuxième lavage plus stringent est parfois ensuite réalisé à l'aide d'un tampon T2 identique au tampon T1 mais avec une plus forte concentration en sel permettant de décrocher les protéines et les acides nucléiques fixés de façon aspécifique sur la colonne ou sur les protéines retenues. L'élution de la protéine est ensuite réalisée grâce à un gradient de tampon T1 avec le tampon T3 jusqu'à atteindre 100 % V/V de tampon T3 contenant de l'imidazole libre (1 M) ou de l'EDTA (100 mM) selon le mode d'élution choisi. Les fractions issues de la purification sont ensuite analysées sur un gel de polyacrylamide en conditions dénaturantes (SDS-PAGE).

6.2.2. Chromatographie d'exclusion de taille

Principe : Cette technique utilisée aussi bien à des fins analytiques que préparatives, permet la séparation des macromolécules en fonction de leurs masses moléculaires et de leurs formes. Des granules de gel poreux sont utilisées. Les molécules de hautes masses moléculaires, dont les diamètres sont supérieurs à celui des pores, sont exclues et sont donc éluées les premières au niveau du volume mort ou V_m . Les molécules de plus faibles masses moléculaires sont éluées plus tardivement car elles passent par les pores des gels. Les solutés sont donc élués dans l'ordre inverse des masses moléculaires et il existe une relation linéaire entre le volume d'élution et le logarithme de la masse moléculaire, pour une certaine gamme de masses moléculaires selon les colonnes, permettant d'estimer après calibration avec des macromolécules de masses moléculaires connues, l'état oligomérique des protéines analysées et de détecter la présence d'éventuels agrégats.

Protocole : Les colonnes qui sont utilisées sont de type Sephadex. Elles sont constituées de polyosides bactériens de type dextran et d'agarose. Une colonne Superdex 200 HR 10/30 (*GE Healthcare*) adaptée pour une séparation optimale dans la gamme 10-600 kDa et couplée à un système FPLC, est lavée en eau puis équilibrée dans un tampon de composition variable selon la protéine analysée. Une colonne Superdex 75 HR 10/30 (*GE Healthcare*) adaptée pour une séparation optimale dans la gamme 3-70 kDa peut également être utilisée. Les échantillons protéiques sont préparés dans le plus petit volume manipulable possible afin d'améliorer la résolution de séparation et sont injectés dans la colonne avec un débit de 0,250 ml/min. La sortie de colonne des protéines est suivie par lecture de l'absorbance pour une longueur d'onde de 280 nm et les valeurs des volumes d'élution pour chaque pic sont relevées et comparées avec celles des standards de masses moléculaires connues.

7. Traitement des protéines purifiées

7.1. Traitements enzymatiques

7.1.1. Clivage par la thrombine

Principe : Suite à la purification et pour certaines protéines étudiées, une étape supplémentaire est nécessaire et consiste à éliminer l'étiquette poly-histidines de la protéine purifiée. Dans certains cas, un site de clivage à la thrombine est présent entre la protéine d'intérêt et l'étiquette. La thrombine est une endoprotéase de la famille des sérines protéases et dont l'activité de clivage est maximale entre un résidu arginine et un résidu P1' au sein d'une séquence P4-P3-Pro-Arg-P1'-P2' où P4 et P3 sont des acides aminés hydrophobes et P1' et P2' des acides aminés non acides ou entre un résidu arginine et un résidu glycine au sein d'une séquence Gly-Arg-Gly.

Protocole : Pour cliver l'étiquette de purification, la protéine d'intérêt est mise en présence de thrombine (*Amersham Biosciences*) à 4 °C sur la nuit et sous agitation douce à raison de cinq unités de thrombine par mg de protéine à digérer. L'efficacité du clivage est ensuite contrôlée par analyse sur gel de polyacrylamide en conditions dénaturantes (SDS-PAGE). L'étiquette clivée est éliminée par passage de la solution sur une colonne de nickel.

7.1.2. Clivage par la protéase du TEV

Principe : Dans certains cas, un site de clivage par la protéase du virus de la gravure du tabac (*Tobacco Etch Virus* ou TEV) est présent entre la protéine d'intérêt et l'étiquette poly-histidines. Cette protéase possède une stabilité et une spécificité de séquence plus élevées que la thrombine et présente également l'avantage de pouvoir facilement être surproduite au sein du laboratoire et en grande quantité. La protéase du TEV est une sérine protéase dont l'activité est maximale sur les épitopes linéaires ENLYFQG et ENLYFQS avec le clivage ayant lieu entre les résidus Q et G ou Q et S.

Protocole : Pour cliver l'étiquette de purification, la protéine d'intérêt est mise en présence de la protéase du TEV surproduite au sein du laboratoire et sous agitation douce. Différents rapports massiques protéase du TEV/protéine cible, temps d'incubation et températures d'incubation ont été testés. L'efficacité du clivage est ensuite contrôlée par analyse sur gel de polyacrylamide en conditions dénaturantes (SDS-PAGE). L'étiquette de purification clivée, ainsi que la protéase du TEV qui possède une étiquette poly-histidines, peuvent être éliminées par une étape supplémentaire de chromatographie d'affinité pour le nickel.

7.2. Concentration des protéines

Principe : Certaines applications nécessitent la concentration des protéines d'intérêt contenues dans un échantillon. La concentration des protéines étudiées dans ce travail est réalisée par l'utilisation d'un système d'ultrafiltration. Une membrane semi-perméable peut séparer les composants d'un mélange selon leurs masses moléculaires. Le seuil de coupure de la membrane est choisi de sorte que la protéine reste dans le rétentat permettant la concentration de la protéine

d'intérêt. Différents seuils de coupure ainsi que différentes natures de membranes sont choisis selon les protéines étudiées.

Protocole : Les concentrateurs qui sont utilisés dans ce travail sont les concentrateurs possédant une membrane en *Poly Ether Sulfone* ou PES (*Vivaspin*) et les concentrateurs possédant une membrane en cellulose régénérée (*Amicon*). La solution protéique à concentrer est déposée sur la membrane préalablement activée avec le tampon puis, les systèmes d'ultrafiltration sont centrifugés à 4 000 g et à 4 °C jusqu'à l'obtention d'une concentration en protéine désirée. Un rotor à godet oscillant en remplacement du rotor fixe est parfois utilisé afin de limiter le phénomène de colmatage de la membrane au cours de la concentration.

7.3. Echanges de tampons

Principe : Certaines applications nécessitent de modifier la composition de la solution tampon d'un échantillon contenant une protéine d'intérêt. Comme alternative à la dialyse, le système qui est utilisé dans ce travail pour les échanges de tampons est basé sur le principe de la chromatographie d'exclusion de taille ou filtration sur gel décrit précédemment.

Protocole : Pour les échanges de tampons, une colonne de type Sephadex est utilisée. Une colonne HiPrep 26/10 (*GE Healthcare*) adaptée pour séparer les macromolécules de plus de 5 kDa des plus petites molécules de moins de 1 kDa et couplée à un système FPLC, est lavée en eau puis équilibrée dans la solution tampon finale désirée. L'échantillon protéique d'un volume maximal de 15 ml est injecté dans la colonne avec un débit de 3 ml/min. Les macromolécules de tailles supérieures à 5 kDa sont éluées en premier dans la solution tampon finale alors que les molécules de tailles inférieures à 1 kDa sont éluées en dernier. La sortie de colonne des protéines est suivie par lecture de l'absorbance pour des longueurs d'ondes de 280 et 260 nm, et la sortie d'autres molécules comme les sels est suivie par lecture de la conductance de la solution. Les fractions récupérées peuvent ensuite être analysées sur un gel de polyacrylamide en conditions dénaturantes (SDS-PAGE). Facultativement, le système d'ultrafiltration précédemment décrit et également utilisé pour la concentration des protéines est parfois utilisé pour les échanges de tampons.

II. TECHNIQUES ANALYTIQUES

1. Contrôle de la qualité et de la quantité des ADN

1.1. Spectroscopie ultraviolet-visible

Principe : La spectroscopie ultraviolet-visible est utilisée pour contrôler la quantité et la qualité des ADN utilisés au cours de ce travail. Les acides nucléiques absorbent la lumière dans l'ultraviolet, dans la gamme de longueur d'onde de 240 à 280 nm. L'absorption de la lumière par les acides nucléiques résulte de la nature aromatique des bases et présente un maximum pour la longueur d'onde de 260 nm.

Lorsque la lumière passe à travers une solution contenant une substance absorbante, son intensité décroît exponentiellement selon la loi de Beer-Lambert, où I_0 est l'intensité de la lumière incidente, I l'intensité de la lumière transmise, l la longueur du trajet optique de la solution traversée exprimée en cm, ϵ le coefficient d'extinction molaire exprimé en $\text{l.mol}^{-1}.\text{cm}^{-1}$ et c la concentration de l'analyte exprimée en mol.l^{-1} :

$$I = I_0 \cdot e^{-\epsilon_{\lambda} l c}$$

Le ratio I/I_0 représente la transmittance de la solution et l'absorbance A_{λ} (valeur sans unité) est reliée au logarithme de ce ratio nous permettant d'écrire :

$$A_{\lambda} = -\log_{10} (I / I_0) = \epsilon_{\lambda} \cdot l \cdot c$$

Par la mesure de l'absorbance à une longueur d'onde donnée d'un échantillon contenant une substance absorbante avec un coefficient d'extinction molaire connu pour cette longueur d'onde et en connaissant la longueur du trajet optique, il est donc possible de déterminer la valeur de c correspondant à la concentration de cette substance dans la solution. Pour les acides nucléiques, il est généralement accepté que les valeurs de coefficients d'extinctions molaires pour la longueur d'onde de 260 nm correspondent à 50, 33 et 40 $\text{ng.cm.}\mu\text{l}^{-1}$ respectivement pour les ADN doubles brins, simples brins et l'ARN permettant de déterminer par mesure de l'absorbance à 260 nm et selon la nature des acides nucléiques leurs concentrations en $\text{ng.}\mu\text{l}^{-1}$.

La pureté des solutions d'acides nucléiques peut être estimée par la mesure des rapports d'absorbance à 260 et 280 nm (A_{260}/A_{280}) qui doit être de l'ordre de 1,8 pour un ADN double brin et de 2,0 pour un ARN (ces valeurs pouvant toutefois dépendre du pH et de la force ionique de la solution). Les protéines présentent un maximum d'absorption à la longueur d'onde de 280 nm. La contamination d'un échantillon d'acides nucléiques par des protéines se caractérise donc par un rapport A_{260}/A_{280} inférieur à 1,8 et la contamination d'un échantillon d'ADN par l'ARN se caractérise par un rapport A_{260}/A_{280} supérieur à 1,8. Le rapport A_{260}/A_{230} peut également être utilisé pour déterminer la qualité des échantillons et sa valeur est souvent supérieure à celle du rapport A_{260}/A_{280} pour une solution pure d'acides nucléiques. La valeur de ce rapport est normalement comprise entre 2,0 et 2,2 et est plus faible dans le cadre d'une contamination par des substances absorbant à 230 nm (EDTA, sucres, phénol, chlorure de guanidinium).

Protocole : La mesure des absorbances des solutions d'acides nucléiques est réalisée à l'aide d'un spectrophotomètre Nanodrop 2 000 (*Thermo Scientific*) qui permet de réaliser les mesures sur des échantillons d'acides nucléiques de 1 à 2 μl (trajet optique de 1 mm). Les valeurs des rapports A_{260}/A_{280} et A_{260}/A_{230} automatiquement calculées permettent de contrôler la qualité des échantillons.

1.2. Electrophorèse sur gel d'agarose

Principe : L'électrophorèse sur gel d'agarose (polymère à base d'agar purifié) est une méthode utilisée pour séparer l'ADN, l'ARN ou les protéines en fonction de leur conformation, de leur poids moléculaire et de leur charge. Cette technique est utilisée pour l'analyse et la séparation des acides nucléiques. Elle est basée sur la séparation de la cathode vers l'anode des acides nucléiques chargés négativement dans un milieu basique, sous l'effet de l'application d'un champ électrique. La migration sous l'effet du champ électrique se réalise dans l'épaisseur du gel poreux d'agarose dans lequel l'échantillon a été déposé. Les pores de ce gel se comportent comme un tamis qui freine le déplacement des macromolécules. Les molécules de plus petites tailles se déplacent plus rapidement et migrent plus loin que les molécules de tailles supérieures. L'augmentation de la

concentration d'agarose dans un gel réduit la vitesse de migration et permet la séparation de fragment d'ADN de plus petite taille. La méthode de révélation des acides nucléiques la plus utilisée est la révélation au bromure d'éthidium ou BET. Lorsqu'il est exposé à des rayonnements ultraviolets, le BET devient fluorescent avec une couleur rouge-orangée qui est vingt fois plus intense lorsqu'il est lié aux acides nucléiques.

Protocole : Les fragments d'acides nucléiques sont analysés par électrophorèse sur gel 1 % W/V agarose suivi d'une révélation par le bromure d'éthidium. Pour un gel 1 % W/V agarose de 50 ml, 0,5 g d'agarose est pesé puis 50 ml d'une solution de 1 X TAE (40 mM Tris pH 8,3, 20 mM acétate et 1 mM EDTA, *Euromedex*) est ajoutée et le mélange est chauffé à l'aide d'un micro-onde jusqu'à l'obtention d'une solution homogène. Le mélange est ensuite refroidi et 2 µl de 10 mg/ml BET (*Invitrogen*) est ajouté. Le gel est laissé à température ambiante jusqu'à polymérisation complète avant d'être utilisé. Les échantillons d'acides nucléiques sont ensuite mélangés avec un tampon de charge 6 X (30 % V/V glycérol, 0,25 % W/V de bleu de bromophénol) utilisé à une concentration finale de 1 X et déposés sur le gel d'agarose, le bleu de bromophénol servant de front de migration. La migration est réalisée à 50 ou 100 V pendant trente à soixante minutes dans du tampon de migration 1 X TAE puis le gel exposé aux ultraviolets.

1.3. Séquençage de l'ADN

Principe : Le séquençage des ADN plasmidiques d'intérêt est réalisé en utilisant la méthode enzymatique de séquençage ou méthode de F. Sanger (*Sanger et al.* 1977). Dans cette méthode, la polymérisation de l'ADN d'intérêt est initiée par une amorce dite de séquençage et complémentaire à une partie du fragment d'ADN à séquencer. L'élongation de l'amorce est réalisée par une enzyme, généralement la séquenase, qui est une enzyme à activité ADN polymérase I et dépourvue d'activité exonucléasique. Les quatre désoxynucléotides (dATP, dCTP, dGTP et dTTP) sont ajoutés ainsi que quatre didésoxynucléosides (ddATP, ddCTP, ddGTP et ddTTP) en plus faibles concentrations et marqués par des fluorochromes différents. Au cours de la polymérisation de l'ADN d'intérêt, ces didésoxynucléosides vont pouvoir être incorporés à la nouvelle chaîne synthétisée empêchant la poursuite de l'élongation. Les acides nucléiques amplifiés sont séparés par une électrophorèse sur capillaire et quatre capteurs (un par longueur d'onde) enregistrent la lumière émise par les fragments terminés par un ddNTP marqué. Ces quatre types de signaux sont superposés et un logiciel traduit les données résultantes en séquence. Par cette méthode, il est ainsi possible de déterminer la séquence d'un fragment d'environ 750 nucléotides.

Protocole : Les ADN plasmidiques d'intérêt et purifiés sont séquencés par la méthode de F. Sanger par le service de séquençage GATC (www.gatc-biotech.com).

2. Contrôle de la qualité et de la quantité des protéines

2.1. Analyses spectroscopiques

2.1.1. Spectroscopie ultraviolet-visible

Principe : La spectroscopie ultraviolet-visible est utilisée pour contrôler la quantité et la qualité des protéines utilisées au cours de ce travail. Les protéines absorbent la lumière dans la région de l'ultraviolet (175 à 350 nm) du fait des groupements carbonyles des liaisons peptidiques (absorption de 190 à 230 nm), des acides aminés aromatiques (absorption de 250 à 320 nm) et des ponts disulfures (absorption de 250 à 300 nm). Comme décrit précédemment pour le dosage des acides nucléiques et selon la loi de Beer-Lambert, la mesure de l'absorbance à une longueur d'onde donnée d'un échantillon, contenant une substance absorbante avec un coefficient d'extinction molaire connu pour cette longueur d'onde, permet de déterminer la concentration de cette substance dans la solution.

Il est considéré que l'absorption des protéines au-delà de 230 nm résulte majoritairement de la contribution des absorptions liées aux résidus aromatiques tryptophanes, tyrosines et phénylalanine ainsi que des ponts disulfures (cystines). L'absorption de la lumière par les protéines est maximale par les résidus tryptophanes et tyrosines et présente généralement un maximum pour la longueur d'onde de 280 nm. Il est accepté que les valeurs de coefficients d'extinctions molaires pour la longueur d'onde de 280 nm correspondent à 5 500, 1 490 et 125 M⁻¹.cm⁻¹ respectivement pour les résidus tryptophanes, tyrosines et pour les cystines. Il a ainsi pu être montré que le coefficient d'extinction molaire $\epsilon_{\text{molaire}}$ (en M⁻¹.cm⁻¹) d'une protéine correctement repliée ou dénaturée, est sensiblement la même et peut être approximée en connaissant les nombres n_w , n_y et n_c respectivement de résidus tryptophanes, tyrosines et cystéines (en tenant compte ou non qu'elles soient oxydées) par l'utilisation de la formule suivante (Pace *et al.* 1995) :

$$\epsilon_{\text{molaire}} = (n_w \cdot 5\,500) + (n_y \cdot 1\,490) + (n_c \cdot 125)$$

Par la mesure de l'absorbance à la longueur d'onde de 280 nm d'un échantillon d'une protéine purifiée de coefficient d'extinction molaire connu et calculé par la précédente formule et en connaissant la longueur du trajet optique, il est donc possible de déterminer la valeur de c correspondant à la concentration (en M) de cette protéine dans la solution.

La valeur d'absorptivité molaire $\epsilon_{\text{molaire}}$ peut être convertie en valeur d'absorptivité massique $\epsilon_{1\%}$ pour une solution à 1 % (en [g/100 ml]⁻¹.cm⁻¹) en connaissant la masse moléculaire M de la protéine :

$$\epsilon_{\text{molaire}} \cdot 10 = \epsilon_{1\%} \cdot M$$

La concentration c de la protéine (en mg.ml⁻¹) peut ainsi être directement obtenue par l'application de la formule suivante :

$$(A_{280} / \epsilon_{1\%}) \cdot 10 = c$$

La qualité et la pureté des solutions protéiques peuvent être évaluées par la mesure du rapport d'absorbance à 260 et 280 nm (A_{260}/A_{280}). En effet, les acides nucléiques absorbent fortement la lumière à 260 nm alors que les protéines présentent un maximum d'absorption à la longueur d'onde de 280 nm. La contamination d'un échantillon protéique par des acides nucléiques

se caractérise donc par un rapport A_{260}/A_{280} généralement supérieur à 0,6. L'agrégation des protéines peut également être détectée par spectroscopie ultraviolet-visible dans la gamme de longueur d'onde 300-500 nm, les particules agrégées étant capable de diffuser la lumière.

Protocole : La mesure des absorbances des solutions protéiques est réalisée à l'aide d'un spectrophotomètre Nanodrop 2 000 (*Thermo Scientific*) qui permet de réaliser les mesures sur des échantillons protéiques de 2 μ l (trajet optique de 1 mm). La valeur du rapport A_{260}/A_{280} automatiquement calculé permet de contrôler la qualité des échantillons.

2.1.2. Spectroscopie à corrélation de photons

Principe : La spectroscopie à corrélation de photons, également connue sous le nom de diffusion dynamique de la lumière ou DLS (pour *Dynamic Light Scattering*), est une technique permettant d'estimer le diamètre hydrodynamique de particules à partir de la mesure de leurs coefficients de diffusion, mais également de caractériser la dispersité d'un échantillon. Elle est basée sur la mesure du mouvement Brownien en milieu liquide. Le mouvement Brownien est le mouvement aléatoire des particules dû aux collisions de ces dernières avec les molécules du solvant. Plus une particule est grosse, plus le mouvement brownien est lent, alors que les particules de petites tailles ont un mouvement brownien beaucoup plus important. La diffusion à 90 ° de la lumière par des particules illuminées à l'aide d'un laser est détectée et présente des fluctuations avec le temps, qui peuvent être reliées au coefficient de diffusion des particules en solution. Ce coefficient de diffusion est relié au rayon hydrodynamique de la particule par l'équation de Stokes-Einstein avec D le coefficient de diffusion des particules en solution, K_B la constante de Boltzmann, T la température absolue, η la viscosité du milieu et r le rayon de la particule sphérique :

$$D = K_B \cdot T / 6 \cdot \pi \cdot \eta \cdot r$$

En connaissant les variables température et viscosité du milieu, il est donc possible d'estimer à partir des coefficients de diffusion des particules mesurés par DLS, les rayons et donc les diamètres hydrodynamiques théoriques de ces particules. Le diamètre hydrodynamique calculé correspond au diamètre hydrodynamique d'une sphère qui possède le même coefficient de diffusion que la particule et peut être relié à la masse moléculaire théorique de la particule. Toutefois, le coefficient de diffusion ne dépend pas uniquement du diamètre hydrodynamique de la particule mais également de la nature de la surface de la particule ainsi que de sa concentration et du type d'ions présent dans la solution. De plus, l'estimation du diamètre hydrodynamique des particules ne peut s'appliquer à la particule que si cette dernière est sphérique et doit être prise avec précaution pour les particules non sphériques.

Protocole : Les expériences de DLS sont réalisées à 8 ou 25 °C avec des solutions protéiques à différentes concentrations selon les cas et après centrifugation pendant dix minutes à 10 000 g et à 4 °C pour éliminer le maximum d'agrégats et impuretés éventuelles, sur un appareil Zetasizer Nano S ZEN 1 600 (*Malvern Instruments*) équipé de cuves en quartz d'une contenance de 70 μ l ou de cuves stériles en plastique d'une contenance de 70 à 100 μ l (*Brand*). Une droite d'étalonnage, représentant les diamètres hydrodynamiques obtenus pour des standards protéiques de masses moléculaires connues, permet d'estimer l'état d'oligomérisation des protéines en solution.

2.1.3. Spectrométrie de masse

Principe : La spectrométrie de masse est une technique d'analyse biophysique permettant de détecter puis d'identifier des molécules d'intérêt par mesure de leurs masses. Son principe réside sur la séparation en phase gazeuse de molécules chargées (ionisées) en fonction de leurs rapports masse/charge.

Le spectromètre de masse comporte une source d'ionisation suivie d'un ou plusieurs analyseurs qui séparent les ions produits selon leur rapport masse/charge ou rapport m/z , et les comptent. Le résultat obtenu est un spectre de masse représentant les rapports m/z des ions détectés et l'abondance relative de ces derniers.

Le spectromètre de masse se compose donc de quatre parties :

- Le système d'introduction de l'échantillon : l'échantillon peut être introduit directement dans la source, sous forme liquide ou solide, ou encore par l'association à une méthode séparative.
- La source d'ionisation va vaporiser les molécules et les ioniser selon un mode choisi.
- L'analyseur permet de séparer les ions en fonction de leur rapport m/z . Il existe des analyseurs permettant de mesurer la masse exacte des analytes. On parle dans ce cas d'analyseurs à temps de vol ou *Time Of Flight* (TOF). Ces analyseurs peuvent être couplés entre eux pour réaliser des expériences de spectrométrie de masse en tandem ou MS/MS.
- Le détecteur et le système de traitement : le détecteur transforme les ions en signal électrique. Plus les ions sont nombreux, plus le courant est important. De plus, le détecteur amplifie le signal obtenu pour qu'il puisse être traité par informatique.

Protocole : Toutes ces expériences ont été réalisées en collaboration avec le service de spectrométrie de masse de la plateforme technique « Centre Commun de Microanalyse des Protéines » ou CCMP, localisée à l'IBCP. Les échantillons à analyser ont pu être fournis sous deux formes : en solution ou sous forme solide dans un gel d'acrylamide. Cette dernière forme nécessite une digestion préalable par la trypsine pour obtenir des fragments protéiques de plus petites tailles qui pourront être analysés.

2.2. Analyses par électrophorèse sur gel d'acrylamide

2.2.1. En conditions dénaturantes

Principe : L'électrophorèse permet d'identifier les protéines selon leurs poids moléculaires par migration au travers d'un gel d'acrylamide sous l'effet d'un champ électrique en présence du détergent SDS. On parle alors de SDS-PAGE ou *SDS-PolyAcrylamide Gel Electrophoresis*. Cette technique a été décrite pour la première fois par L. Ornstein et B. J. Davies (Ornstein 1964) puis modifiée par U. K. Laemmli avec l'incorporation de SDS (Laemmli 1970). Le réseau de polyacrylamide est formé par co-polymérisation de monomères d'acrylamide (ou acryl) en présence de petites quantités de bis-acrylamide (NN'-méthylène-bisacrylamide), la réaction étant catalysée par l'ajout d'ammonium persulfate ou APS, ainsi que de N, N, N', N'-tétraméthyléthylènediamine ou TEMED. Le TEMED catalyse la décomposition des ions persulfates pour donner des radicaux libres nécessaires à la réaction. La grosseur des pores du gel est fonction de la concentration d'acrylamide et du taux de

bis-acrylamide. Plus la concentration est élevée, plus les pores seront petits. Le système utilisé comprend deux modes successifs de migration électrophorétique, une migration dans un gel de concentration, puis une migration dans un gel de séparation. Le passage de l'une à l'autre se fait avec un changement de pH et par une augmentation de concentration du gel en acrylamide. L'objectif du gel d'électrophorèse en deux parties, ou électrophorèse discontinue, est de pallier aux problèmes liés aux faibles quantités de protéines et aux volumes parfois non négligeables déposés dans les puits. On réalise donc une étape de concentration des échantillons jusqu'à l'obtention de zones protéiques très fines dans le gel de concentration ou *stacking gel*, qui migreront ensuite dans le gel de séparation ou *resolving gel*. Le SDS est un détergent qui interagit avec les protéines. Ainsi, il dénature et confère une charge globale négative aux protéines. De cette façon, seule la masse moléculaire des protéines sera le facteur de leur séparation. Ceci permet la migration des protéines dans la matrice à l'aide d'un courant électrique (de la cathode vers l'anode) et la séparation des protéines s'effectue uniquement en fonction de leurs masses moléculaires (les protéines ayant une masse moléculaire plus faible migreront plus rapidement).

Protocole : Différents pourcentages de gels ont pu être utilisés pour les protéines étudiées dans ce travail (**tableau 1**) :

	GEL DE SÉPARATION				GEL DE CONCENTRATION	
	8 %	10 %	12 %	15 %		4 %
30 % Bis-Acrylamide	2,6 ml	3,3 ml	4 ml	5 ml	30 % Bis-Acrylamide	660 µl
3 M Tris HCl pH 8,8	1,25 ml	1,25 ml	1,25 ml	1,25 ml	1 M Tris HCl pH 6,8	625 µl
H ₂ O	5,6 ml	5 ml	4,24 ml	3,24 ml	H ₂ O	3,48 ml
10 % W/V SDS	400 µl	400 µl	400 µl	400 µl	10 % W/V SDS	200 µl
10 % W/V APS	100 µl	100 µl	100 µl	100 µl	10 % W/V APS	50 µl
TEMED	10 µl	10 µl	10 µl	10 µl	TEMED	50 µl

Tableau 1 : Composition des gels de séparation et de concentration.

Les échantillons à analyser sont mélangés au tampon de charge pour gel dénaturant cinq fois concentré (**tableau 2**) puis chauffés cinq à dix minutes à 95-100 °C avant d'être déposés sur le gel. Pour certaines protéines, les échantillons ne sont pas chauffés avant d'être déposés et un tampon Laemmli dépourvu d'agent réducteur peut être utilisé. Une solution de marqueurs de masses moléculaires connues (*Prestained Protein Ladder, Euromedex*) est également déposée afin de pouvoir estimer la masse moléculaire des protéines d'intérêt. La migration est réalisée à 90 V (voltage constant) pendant près de dix minutes (gel de concentration) puis à 100-140 V (voltage constant) pendant près de 1 h 30 dans le tampon d'électrophorèse Tris-Glycine-SDS (25 mM Tris-base, 192 mM glycine, 0,1 % W/V SDS, *Euromedex*).

TAMPON LAEMMLI POUR GEL DÉNATURANT (5 X)	
Tris HCl pH 6,8	0,25 M
SDS	5 % W/V
Bleu de bromophénol	0,5 % W/V
β-mercaptoéthanol	0,7 M
Glycérol	20 % V/V

Tableau 2 : Composition du tampon de charge 5 X pour gel dénaturant.

2.2.2. En conditions natives

Principe : Dans l'électrophorèse en conditions non dénaturantes, les molécules sont séparées dans leur état le plus proche possible de leur état natif. Les étapes préalables, permettant l'obtention de la solution contenant les molécules à séparer, doivent éviter les conditions dénaturantes. Les analyses se font donc, à la différence de l'électrophorèse en conditions dénaturantes, en l'absence de SDS et les échantillons ne sont pas chauffés avant les dépôts. Dans ce système, la séparation des macromolécules se fait en fonction de nombreux critères. Pour un support et un voltage donnés, la vitesse de migration dépendra donc de :

- la charge « native » de la molécule, qui dépend du pH de la solution tampon utilisée pour la migration. Selon le signe de la charge globale de la macromolécule, ces dernières pourront migrer vers l'anode (molécules chargées négativement) ou vers la cathode (molécules chargées positivement). Ainsi, plus la charge globale de la macromolécule sera importante, plus les molécules migreront vite.
- la masse moléculaire. Plus la masse moléculaire est importante, plus la vitesse de migration de la molécule est faible.
- la conformation tridimensionnelle. A masse moléculaire et charges égales, une molécule pourra selon sa forme être plus ou moins ralentie dans sa progression par le support. Par exemple, une molécule de structure globulaire sera moins ralentie qu'une molécule de structure fibreuse.

Cette technique présente certains avantages comme celui de permettre la détermination de l'état oligomérique (monomères, dimères) et conformationnel (protéines dénaturées ou correctement repliées) des protéines analysées. Elle peut également permettre sous certaines conditions de suivre des phénomènes de liaisons, qui peuvent entraîner soit un changement conformationnel (oligomérisation) soit le masquage de charges à la surface d'une protéine cible par l'ajout d'un partenaire protéique ou d'un ligand.

Protocole : Le système qui est utilisé pour la réalisation des électrophorèses en conditions natives, est basé sur celui développé par T. McLellan (McLellan 1982) qui utilise des gels continus (sans gel de concentration) et des solutions tampons adaptées pour des analyses dans une gamme de pH de 3,8 à 10,2. Les gels réalisés dans ce travail possèdent un taux de réticulation de 8 % (pour 10 ml : 5,33 ml d'eau, 2,67 ml de 30 % Bis-acrylamide, 2 ml de 5 X tampon d'électrophorèse, 60 µl de 10 % W/V APS et 10 µl de TEMED). Un tampon d'électrophorèse 1 X permettant l'obtention d'un pH de 7,4 au cours de la migration est utilisé (5 X tampon d'électrophorèse pH 7,4 : 1,464 % W/V imidazole, 4,17 % W/V Hepes). Les échantillons à analyser sont mélangés au tampon de charge pour gel natif deux fois concentré (pour 2 X : 30 % V/V glycérol, 0,1 X tampon d'électrophorèse) avant d'être déposés sur le gel. Une première migration est réalisée avant les dépôts à 100 V (voltage constant) pendant trente minutes. Les échantillons sont ensuite déposés puis la migration réalisée à 100-140 V (voltage constant) pendant 1 h à 4 °C.

2.2.3. Méthodes de révélations

2.2.3.A. Bleu de coomassie colloïdal

Principe : Le bleu de coomassie colloïdal se lie aux acides aminés basiques et aromatiques des protéines et permet un seuil de détection de 5 ng. Cette coloration présente l'avantage d'être compatible avec les expériences de spectrométrie de masse pour les analyses de protéines excisées à partir de gels de polyacrylamide.

Protocole : Une solution commerciale de bleu de coomassie colloïdal (*Page Blue Protein Staining Solution, Fermentas*) a été utilisée pour ce mode de coloration.

2.2.3.B. Nitrate d'argent

Principe : La coloration au nitrate d'argent, d'une plus grande sensibilité que la coloration au bleu de coomassie, permet de vérifier la pureté de l'échantillon (seuil de détection de 0,1 ng). Pour la coloration au nitrate d'argent, les groupements sulfhydryles SH et carboxyliques COOH des chaînes latérales des protéines peuvent réduire l'ion argent. L'ion argent réduit est alors sensible à la lumière. Cette réaction peut être arrêtée en milieu acide.

Protocole : Le gel est fixé dans une solution de 50 ml de 40 % V/V éthanol en présence de 25 µl de 37 % W/V formaldéhyde pendant dix minutes. Le gel est ensuite rincé deux fois à l'eau distillée pendant cinq minutes puis incubé dans une solution de $\text{Na}_2\text{S}_2\text{O}_3$ (0,02 % W/V dans l'eau) pendant dix minutes. Le gel est rincé dans de l'eau pendant vingt secondes puis est mis en contact avec le nitrate d'argent (0,1 % W/V dans de l'eau) pendant dix minutes. La révélation a lieu après un rinçage rapide dans l'eau dans une solution de Na_2CO_3 (3 % W/V), $\text{Na}_2\text{S}_2\text{O}_3$ (0,001 % W/V) en présence de 25 µl de 37 % formaldéhyde W/V pendant dix minutes. La coloration est arrêtée par ajout d'acide citrique 2,3 M ou d'acide acétique 5 % W/V.

2.2.3.C. Immuno-empreinte

Principe : L'immuno-empreinte est une méthode permettant la détection et l'identification de protéines spécifiques dans un échantillon biologique à l'aide d'anticorps spécifiques.

Protocole : Après électrophorèse sur gel de polyacrylamide, une membrane de nitrocellulose (Hybond-CExtra, *GE Healthcare*) est placée en face du gel et un transfert des protéines est effectué sous l'effet d'un courant électrique (électrotransfert), dans du tampon de transfert (0,1 % W/V SDS, 3 mg/ml final Tris-base, 20 % V/V) pendant 1 h à 100 V à ampérage constant (environ 350 mA). Les protéines migrent depuis le gel vers la membrane en conservant leurs positions relatives. Les protéines se fixent à la membrane grâce à des interactions hydrophobes et ioniques. Afin de minimiser les interactions non spécifiques entre les anticorps et la membrane de nitrocellulose, une étape de saturation est réalisée en plongeant la membrane de nitrocellulose dans une solution de TNT (10 mM Tris pH 7,5, 150 mM NaCl, 0,1 % W/V Tween-20) puis dans une solution de tris-lait (5 g de lait écrémé en poudre, 100 ml TNT) sur la nuit et sous agitation à 4 °C. L'anticorps primaire de

souris (*Abcam*) dilué au 1/1000ème dans du TNT (0,5 µg/ml final) et reconnaissant spécifiquement l'étiquette poly-histidines, est alors ajouté pendant une heure et sous agitation. La membrane est rincée trois fois avec du tampon TNT afin d'enlever les anticorps primaires non liés à la membrane. Puis, celle-ci est mise en contact avec un anticorps secondaire (*Sigma*) de chèvre anti-souris couplé à la phosphatase alcaline, et dilué au 1/2 000ème dans du tampon TNT (12,5 ng/ml final) pendant une heure sous agitation. La membrane est rincée plusieurs fois avec du tampon TNT. La révélation a alors lieu par détection colorimétrique en utilisant un substrat chromogénique commercial BCIP-NBT (*Sigma-Fast*). L'enzyme convertit un colorant soluble jaunâtre en une forme insoluble mauve révélant alors la protéine d'intérêt. La coloration est arrêtée par rinçage de la membrane avec de l'eau distillée.

2.3. Analyse par pontage chimique des protéines

Principe : Les pontages covalents permettent de figer de façon permanente une interaction protéine-ligand ou protéine-protéine. Il est ensuite aisé d'observer après migration sur SDS-PAGE de telles interactions. Dans le cas présent, cette technique nous permet d'analyser l'état oligomérique de la protéine en solution. Deux types d'agent pontant ont été utilisés dans ce travail, le Bis(SulfoSuccinimidyl) Suberate ou BS3 et le glutaraldéhyde ou GTA.

2.3.1. Pontage covalent au BisSulfoSuccinimidyl Suberate (BS3)

Protocole : Le Bis SulfoSuccinimidyl Suberate (BS3) est un ponteur homobifonctionnel non clivable qui contient à ses deux extrémités une fonction ester de N-hydroxysulfosuccinimide (NHS) avec un bras espaceur de 11 Å. Le NHS peut réagir avec les amines primaires (par exemple des lysines) pour former une liaison amide stable avec libération du groupe N-hydroxysulfosuccinimide. La solution réactionnelle comprends 2 µl de tampon 1 M HEPES pH 8,0, 2,4 µl d'une solution de 5 M NaCl, 2 µg de protéine, 0,5 µl d'une solution de 80 mM BS3 pour les essais BS3 1 X et 1 µl pour les essais BS3 2 X (*Pierce*), et de l'eau qui est ajoutée pour atteindre un volume final de 20 µl. La solution réactionnelle est incubée pendant trente minutes à température ambiante puis 2,4 µl d'une solution de 250 mM Tris HCl pH 7,5, 30 mM glycine et 0,7 µl d'eau sont ajoutés pour arrêter la réaction. La solution est ensuite incubée quinze minutes à température ambiante. Une analyse par électrophorèse sur gel discontinu de polyacrylamide 8 % en conditions dénaturantes est ensuite réalisée.

2.3.2. Pontage au glutaraldéhyde (GTA)

Protocole : Les groupements amines des protéines peuvent réagir avec des composés bi-aldéhydiques, comme le glutaraldéhyde, pour former des dérivés activés capables de se lier à d'autres protéines. Pour le principe, les fonctions amines du squelette carboné, ou des chaînes latérales de certains acides aminés des protéines, peuvent former des bases de Schiff (fonctions imines) avec les fonctions aldéhydes des deux extrémités du glutaraldéhyde, chaque extrémité liant dans le meilleur cas deux protéines différentes. La solution réactionnelle comprend 5 µl de solution protéique à laquelle est ajouté 1 µl d'une solution de 0,6 % W/V glutaraldéhyde pour les essais

glutaraldéhyde 1 X et 1,2 % W/V pour les essais glutaraldéhyde 2 X (*Pierce*). La solution est mise à incuber vingt minutes à température ambiante avant d'être analysée par électrophorèse sur gel discontinu de polyacrylamide 8 % en conditions dénaturantes.

3. Mesure des interactions protéines/ligands

3.1. Analyse par Titration Calorimétrique Isotherme (ITC)

Principe : La Titration Calorimétrique Isotherme ou ITC (pour *Isothermal Titration Calorimetry*) est une technique analytique qui permet de mesurer l'affinité et d'autres paramètres thermodynamiques concernant la liaison de deux molécules. L'appareil peut mesurer directement et de façon précise la chaleur produite ou absorbée par l'interaction des molécules. La mesure de cette chaleur rend possible la détermination directe, en une expérience, de la constante d'affinité à l'équilibre (K_D), la stœchiométrie de la liaison (n) ainsi que l'enthalpie (ΔH) et l'entropie (ΔS) de la réaction.

Protocole : Les expériences d'ITC ont été réalisées en collaboration avec le service de de la plateforme de biophysique, localisée à l'IBCP. L'appareil utilisé est un calorimètre ITC200 (*Microcal, GE Healthcare*). La cellule de la sonde a un volume de 300 μ l dans laquelle la solution protéique est disposée. La seringue est remplie avec 60 μ l du ligand d'intérêt. Un contrôle est effectué dans lequel le tampon seul est titré avec le ligand d'intérêt, afin de calculer la chaleur de dilution. Pour un premier essai, la concentration en protéine dans la cellule est choisie de façon à ce qu'elle soit dix fois supérieure à la constante de dissociation prédite.

3.2. Analyse par mesure de la fluorescence intrinsèque des protéines

Principe : La fluorescence intrinsèque des protéines, et en particulier la fluorescence du tryptophane, peut être utilisée pour mesurer l'interaction de ligands avec des protéines (Van de Weert & Stella 2011). Cette technique est basée sur la propriété solvatochromique de la fluorescence des tryptophanes des protéines (dépendance de la fluorescence vis-à-vis de l'environnement du fluorophore) permettant de suivre l'évolution de différents paramètres de cette fluorescence (intensité et longueur d'onde d'émission maximale) en fonction d'un paramètre extérieur (pH, température, agent dénaturant, ligand).

Les résidus tryptophanes des protéines participent fréquemment à la fixation de molécules comme la fixation de substrats au sein d'un site actif. La liaison d'un ligand à proximité d'un fluorophore comme le tryptophane, va généralement pouvoir altérer ses propriétés de fluorescence au niveau de l'intensité d'émission, mais également de la longueur d'onde d'émission maximale de fluorescence. Ce changement de fluorescence peut être suivi en fonction de la concentration du ligand ajouté jusqu'à la stabilisation du signal décrivant un phénomène de saturation de la protéine avec le ligand et permettant d'extraire les paramètres thermodynamiques d'interaction comme la constante de dissociation, l'indice de coopérativité (indice de Hill) et parfois la stœchiométrie de liaison.

Un phénomène important doit être pris en compte au cours de ces expériences et constitue le phénomène de filtre-interne. L'effet de filtre interne (ou dispersion optique) désigne l'absorption de la lumière à une longueur d'onde d'excitation ou d'émission choisie d'un composé (comme un ligand) ajouté dans la solution du fluorophore étudié. Si un ligand absorbe à la longueur d'onde d'excitation du fluorophore, le nombre de photons disponibles pour exciter le fluorophore sera réduit, induisant une diminution de son excitation et donc de sa fluorescence. Egalement, si un ligand absorbe à la longueur d'onde d'émission du fluorophore, le nombre de photons issus de l'émission du fluorophore et arrivant au détecteur sera réduit, d'où une apparente diminution de la fluorescence. L'effet de filtre interne peut également être rencontré pour des ligands ne possédant pas de propriétés d'absorption mais possédant des propriétés de diffusion de la lumière. La fluorescence observée doit donc toujours être corrigée pour cet effet au cours des expériences de titration d'une protéine avec un ligand, par l'utilisation d'une solution contenant un analogue du fluorophore en absence de la protéine (le N-acétyl-L-tryptophanamide ou NATA dans le cadre de la fluorescence des tryptophanes).

Un autre phénomène appelé atténuation collisionnelle ou atténuation dynamique peut également se produire. Ce phénomène n'est pas le résultat de la formation d'un complexe stable entre un fluorophore et un ligand (atténuation statique) et peut gêner l'observation de la formation du complexe. L'atténuation collisionnelle résulte d'une interaction transitoire entre un fluorophore excité et un atome ou une molécule facilitant les transitions non-radiatives du fluorophore excité vers son état stable. Dans le cadre de la fluorescence des tryptophanes, l'oxygène, l'iodure ou l'acrylamide constituent des agents atténuateurs dynamiques connus des tryptophanes. L'utilisation d'un contrôle adapté comme le NATA, est nécessaire pour exclure la possibilité d'un phénomène d'atténuation dynamique du fluorophore par le ligand.

Si on considère le cas le plus simple avec la formation d'un complexe 1:1 (ou une classe de sites équivalents) entre une protéine P et un ligand L induisant une atténuation statique de la fluorescence émise par P et en absence de phénomène de coopérativité de liaison, l'intensité de fluorescence émise qui est mesurée, est une simple addition de celles de la protéine libre et complexée, et peut être suivie par l'équation suivante où %_{att} est le pourcentage d'atténuation de l'intensité de fluorescence émise par P à la concentration L_{ajouté} de ligand (M) et corrigée par le NATA, %_{att, max} le pourcentage maximum d'atténuation de l'intensité de fluorescence émise par P et corrigée par le NATA, K_d la constante de dissociation du complexe (M), F_{protéine} et F_{NATA} les intensités de fluorescences émises respectivement par la protéine et le NATA à la concentration L_{ajouté} de ligand (M) et F₀_{protéine} et F₀_{NATA} les intensités de fluorescences initiales (sans ligand) émises respectivement par la protéine et le NATA :

$$\%_{att} = \frac{\%_{att, max} \cdot [L]_{ajouté}}{K_d + [L]_{ajouté}} \quad \%_{att} = \left(1 - \frac{F_{Protéine}/F0_{protéine}}{F_{NATA}/F0_{NATA}} \right) \cdot 100$$

L'équation de gauche permet par régression non linéaire d'obtenir la valeur du K_d (constante de dissociation) de l'interaction (Van de Weert & Stella 2011) et n'est valable que lorsque la concentration en ligand ajoutée est bien supérieure à la concentration en protéine (au moins dix fois supérieure, dans ces conditions, on peut approximer la concentration en ligand libre par la concentration en ligand ajoutée).

Protocole : La spectroscopie de fluorescence est utilisée pour contrôler le changement d'environnement des tryptophanes des protéines par l'ajout d'un ligand et en déduire la constante de dissociation du complexe. Pour cela, des plaques de 96 puits adaptées aux mesures de fluorescence sont utilisées (Microplate, 96 Well, PS, F-bottom, Black, Non-binding, *Greiner Bio-One*). Les mesures sont réalisées à 30 °C avec 200 µl d'échantillon protéique par puits dans un appareil Infinite M1 000 (*Tecan*) (réglage gain 200, fente d'excitation 8 nm, fente d'émission 5 nm). Un blanc avec le tampon seul et titré également avec le ligand est réalisé afin de supprimer le pic Raman de l'eau. Un contrôle avec une solution de NATA, un équivalent du L-tryptophane, est également réalisé dans le même tampon que la protéine et titré avec le ligand afin de corriger les effets de dilution par l'ajout de ligand, l'effet de filtre-interne et vérifier l'absence d'atténuation collisionnelle de fluorescence. La longueur d'onde d'excitation est réglée à 290-295 nm permettant l'excitation quasi sélective des tryptophanes et le spectre d'émission est enregistré de 310 à 370 nm. Les valeurs d'intensité de fluorescence correspondent à la somme des intensités de fluorescence émise, mesurées pour chaque longueur d'onde dans la gamme spectrale de 310 à 370 nm. La détermination des paramètres des équations est possible par une analyse par régression non linéaire des données expérimentales à l'aide du serveur ZunZun (www.zunzun.com).

3.3. Analyse par fluorimétrie différentielle à balayage thermique

Principe : La fluorimétrie différentielle à balayage thermique ou DSF (pour *Differential Scanning Fluorimetry*) est une méthode permettant d'identifier des ligands interagissant *in vitro* avec une protéine d'intérêt et de déterminer la force de cette interaction (Pantoliano *et al.* 2001). Cette méthode permet de suivre la dénaturation d'une protéine induite par une augmentation progressive de la température en utilisant une sonde fluorescente extrinsèque (comme le Sypro Orange) ayant la capacité de se lier aux régions hydrophobes des protéines qui sont exposées au cours de la dénaturation. Ces molécules fluorescentes reportrices ont la particularité de n'être que peu fluorescentes dans un milieu aqueux et d'acquérir une fluorescence élevée en interagissant avec les régions hydrophobes exposées sur les protéines dénaturées. Du fait que la plupart des protéines sont déstabilisées lors d'une augmentation de la température, la liaison des sondes fluorescentes doit également augmenter en fonction de la température, ce qui se traduit par une augmentation de leur fluorescence (**figure 26**). La mesure de la fluorescence d'une sonde en présence de la protéine et en fonction de la température permettra de déterminer les paramètres de thermostabilité de la protéine comme la température de demi-dénaturation ou T_m .

Un partenaire (ligand ou protéine) se liant à une protéine d'intérêt va généralement augmenter sa stabilité et augmenter son T_m (**figure 26**). L'augmentation de cette stabilité peut être quantifiée par la différence de T_m entre la forme complexée et la forme libre (ΔT_m). Il a été montré que cette différence pour une concentration en ligand donnée, pouvait être utilisée pour estimer la constante de dissociation d'un complexe dans des conditions où la protéine est saturée avec ce ligand (concentration de ligand au moins dix fois supérieure au K_d) (Pantoliano *et al.* 2001).

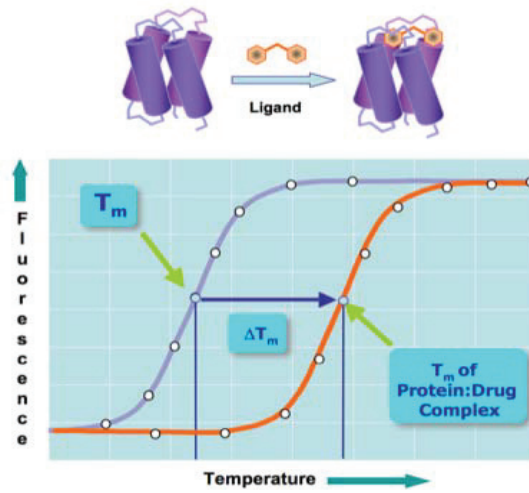


Figure 26 : Principe de la méthode DSF d'après www.thermofluor.org.

Ainsi, il a été montré (Zhang & Monsma 2010) que le suivi indirect de la dénaturation d'une protéine se dénaturant suivant un modèle tout ou rien deux états (replié ou dénaturé), et réversible suit une sigmoïde de Boltzmann et peut être représentée par l'équation suivante où F_{obs} est la fluorescence de la sonde à la température T (Kelvin), T_m la température de demi-dénaturation de la protéine (Kelvin), R la constante des gaz parfaits ($8,3145 \text{ J.mol}^{-1}.\text{K}^{-1}$ ou $1,9872 \text{ cal.mol}^{-1}.\text{K}^{-1}$), M_n et M_u les pentes des portions linéaires de la fluorescence respectivement pré et post-transition, Y_n et Y_u les valeurs d'interceptions sur l'axe des ordonnées des portions linéaires de la fluorescence respectivement pré et post transition, $\Delta H_u^{T_m}$ le changement d'enthalpie de dépliement de la protéine (J.mol^{-1} ou cal.mol^{-1}) à la température T_m et ΔC_{pu} la capacité thermique de dépliement de la protéine ($\text{J.mol}^{-1}.\text{K}^{-1}$ ou $\text{cal.mol}^{-1}.\text{K}^{-1}$) :

$$F_{obs} = \frac{(Y_n + M_n \cdot T) + (Y_u + M_u \cdot T) \cdot \exp \left\{ \frac{-\Delta H_u^{T_m}}{R} \cdot \left(\frac{1}{T} - \frac{1}{T_m} \right) + \frac{\Delta C_{pu}}{R} \cdot \left[\ln \left(\frac{T}{T_m} \right) + \frac{T_m}{T} - 1 \right] \right\}}{1 + \exp \left\{ \frac{-\Delta H_u^{T_m}}{R} \cdot \left(\frac{1}{T} - \frac{1}{T_m} \right) + \frac{\Delta C_{pu}}{R} \cdot \left[\ln \left(\frac{T}{T_m} \right) + \frac{T_m}{T} - 1 \right] \right\}}$$

En absence de ligand, $T_m = T_0$, $\Delta C_{pu} = \Delta C_{pu}^{T_0}$ et $\Delta H_u^{T_m} = \Delta H_u^{T_0}$. En considérant que le ligand n'interagit qu'avec la forme repliée de la protéine et non la forme dénaturée, la constante de dissociation du complexe à la température T_m ($K_{L(T_m)}$) peut être calculée pour une seule concentration en ligand (considérée comme saturante pour la protéine) selon l'équation suivante où $[L_{T_m}]$ est la concentration en ligand libre à la température T_m et qui est approximée par la concentration totale en ligand lorsque la concentration en ligand est au moins dix fois supérieure à la concentration totale en protéine :

$$K_{L(T_m)} = \frac{[L_{T_m}]}{\exp \left\{ \frac{-\Delta H_u^{T_0}}{R} \cdot \left(\frac{1}{T_m} - \frac{1}{T_0} \right) + \frac{\Delta C_{pu}^{T_0}}{R} \cdot \left[\ln \left(\frac{T_m}{T_0} \right) + \frac{T_0}{T_m} - 1 \right] \right\} - 1}$$

Enfin, la constante de dissociation du ligand à une température T ($K_{L(T)}$) peut être estimée par l'équation suivante où $\Delta H_{L(T)}$ représente l'enthalpie de liaison du ligand, qui est le seul paramètre non déterminé dans l'expérience, mais qui peut être raisonnablement approximée par -5 à -10 kcal.mol⁻¹ (Lo *et al.* 2004):

$$K_{L(T)} = \frac{K_{L(T_m)}}{\exp\left\{-\frac{\Delta H_{L(T)}}{R} \cdot \left(\frac{1}{T} - \frac{1}{T_m}\right)\right\}}$$

Protocole : La méthode DSF est utilisée pour suivre la dénaturation des protéines d'intérêt et évaluer l'effet de différents ligands potentiels (fixation ou non) et éventuellement de déterminer les constantes de dissociations des complexes. L'expérience est réalisée dans un appareil RotorGene Q HRM (*Qiagen*) utilisé également pour les expériences de PCR quantitatives. Des tubes adaptés à l'appareil et d'une contenance maximale de 100 µl sont utilisés (*Qiagen*). L'appareil comporte un système permettant un contrôle précis de la température associé à une centrifugation des échantillons au cours de l'expérimentation permettant d'éviter leur évaporation. Un détecteur permet de mesurer la fluorescence émise à une longueur d'onde choisie après excitation à une longueur d'onde adaptée pour la sonde fluorescente.

La sonde fluorescente Sypro Orange (5 000 X initial dans 100 % V/V DMSO, *Sigma*) est utilisée pour nos expériences. Différentes concentrations de protéines et de Sypro Orange sont testées afin d'optimiser le signal de fluorescence, détecté au cours de la dénaturation de la protéine. Les expériences sont réalisées avec ou sans ligand dans un volume final de 20 µl et la gamme de température 35-95 °C a été choisie avec des pas de 1 °C et des pauses de cinq secondes entre chaque degré selon les recommandations du fournisseur de l'appareil. La longueur d'onde d'excitation de la sonde fluorescente est réglée sur 460 nm et celle de l'émission sur 510 nm. Les valeurs de T_m peuvent être déterminées rapidement pour chaque courbe de dénaturation par les dérivées premières de ces courbes. Les paramètres des équations peuvent être déterminés par une analyse par régression non linéaire des données expérimentales à l'aide du serveur ZunZun (www.zunzun.com). Une différence de T_m est généralement considérée comme significative au-delà de 0,5 °C.

4. Analyses structurales

4.1. Analyse par diffusion des rayons X aux petits angles

Principe : La diffusion des rayons X aux petits angles en solution est une méthode d'étude structurale à basse résolution. Elle permet d'obtenir des informations sur la conformation des macromolécules en solution, ainsi que sur la formation d'assemblages supramoléculaires. La diffusion en solution ne nécessite pas de cristal ou de préparation particulière des échantillons, mais nécessite des solutions monodisperses de macromolécules. Des méthodes de modélisation ont été développées, qui permettent de calculer une enveloppe SAXS, équivalente à une structure basse résolution de la macromolécule étudiée. Cette technique permet également de positionner dans les enveloppes SAXS, des structures atomiques de domaines de protéines ou de complexes macromoléculaires obtenus par cristallographie aux rayons X ou RMN. La diffusion en solution n'est donc pas soumise aux principales limitations de la cristallographie (cristallogénèse) et de la RMN

(taille de la molécule étudiée). Elle apparaît comme une technique complémentaire pour l'étude structurale des macromolécules biologiques et l'obtention de modèles tridimensionnels (Putnam *et al.* 2007).

Pour une solution monodisperse idéale, la masse moléculaire et le rayon de giration de la particule en solution peuvent être déterminés à partir de la courbe de diffusion des rayons X aux petits angles. Selon la loi de Guinier, la courbe de diffusion aux petits angles est définie par l'équation suivante, où $I(0)$ est l'intensité des rayons X incidents (valeur non connue à cause du piège à rayons X placé au centre du détecteur), $I(S)$ l'intensité des rayons X diffusés, S le vecteur de diffusion ($S=2\sin\theta/\lambda$ où 2θ est l'angle entre le faisceau de rayons X incident et le faisceau de rayons X diffusés) et R_g le rayon de giration de la particule en solution :

$$I(S) = I(0) \cdot \exp\left(-\frac{S^2 R_g^2}{3}\right)$$

Cette équation peut être transformée pour donner l'équation utilisant les mêmes termes :

$$\ln I(S) = \ln I(0) - \left(\frac{S^2 R_g^2}{3}\right)$$

L'intersection de la droite $\ln I(S) = f(S^2)$ sur l'axe des ordonnées nous permet de déterminer la valeur de $I(0)$ par extrapolation et la pente de cette droite nous permet d'obtenir la valeur du rayon de giration de la particule en solution. Ainsi, la valeur de la masse moléculaire de la particule peut être déterminée grâce à la valeur de l'intensité à l'origine $I(0)$ avec l'équation suivante ou V_{sol} est le volume de la solution, c la concentration de la particule, M la masse moléculaire de la particule, N le nombre d'Avogadro, m_p le nombre d'électrons de la particule, V_p le volume spécifique partiel de la particule et ρ_s la densité électronique du solvant :

$$I(0) = V_{sol} \times \frac{cM}{N} \times \left[\left(\frac{m_p N}{M_w} \right) - \rho_s V_p \right]^2$$

Le calcul de la fonction de distribution des distances, correspondant à la transformée de Fourier inverse des intensités diffusées (réalisable avec le logiciel GNOM (Svergun 1992)), permet de déterminer la valeur de la distance interatomique maximale (D_{max}) de la particule ainsi que son rayon de giration R_g . Ces valeurs constituent le point de départ de la construction d'un modèle *ab initio* dont la courbe théorique se superpose avec la courbe de diffusion expérimentale (réalisable avec le logiciel DAMMIF (Franke & Svergun 2009)).

Protocole : Des mesures de SAXS sont réalisées pour différentes concentrations de protéines ainsi que pour les tampons seuls. Les mesures sont effectuées à l'Institut de Biochimie et Biophysique Moléculaire et Cellulaire (IBBMC, Université de Paris-Sud 11, Orsay) en collaboration avec le Dr Patrice Vachette. L'instrument utilisé pour le SAXS comprend une anode tournante Microstar (Brucker AXS) et un monochromateur pour générer des rayons X à une longueur d'onde de 1,54 Å (raie K α du cuivre). Les rayons sont focalisés sur le porte échantillon par des collimateurs. Un capillaire contenant 15 μ l de solution protéique est installé dans le porte échantillon thermostaté à

15 °C et sous vide. Un détecteur bidimensionnel Vanted 2 000 (*Brucker AXS*) est positionné à 67 cm de l'échantillon. Les images obtenues sont intégrées pour calculer une représentation de l'intensité de diffusion I des rayons X en fonction du vecteur de diffusion S . Le spectre de diffusion obtenu ($\log I=f(S)$) contient des informations à basse résolution (aux petits angles pour S petit) et à haute résolution (aux grands angles pour S grand) sur la structure de la particule en solution.

4.2. Analyse par cristallographie aux rayons X

4.2.1. La cristallogénèse et les méthodes associées

4.2.1.A. Principe général

Principe : La cristallisation est un processus permettant le passage d'une molécule d'un état soluble à un état solide ordonné. Un cristal est un arrangement périodique et ordonné d'un motif composé d'un ensemble de molécules se répétant à l'identique d'une maille à l'autre. Le principe de la cristallisation est d'induire un changement d'état de la protéine de sa phase soluble à une phase solide et ordonnée, ces deux états étant en équilibre. Ce phénomène de transition de phase dépend de nombreux paramètres et peut se diviser en deux étapes : la germination ou plus communément appelée « nucléation » et la croissance cristalline. Pour les protéines, la difficulté consiste à trouver un compromis entre les différents facteurs cinétiques et thermodynamiques qui contrôlent sa solubilité, qui est tributaire des interactions entre les acides aminés de surface et le solvant. Pour induire la nucléation, il faut amener la protéine dans un état soluble instable de sursaturation (état de nucléation) en diminuant lentement sa solubilité (**figure 27**). Ceci revient à déstabiliser les interactions protéines-solvant au profit d'interactions protéines-protéines. Des interactions peuvent conduire à la précipitation (phase solide amorphe) ou au contraire à des interactions spécifiques entre les molécules de protéines conduisant à la formation de germes cristallins (nucléation). Le retour à l'équilibre thermodynamique s'effectue alors plus ou moins rapidement par la croissance de la phase solide (les cristaux dans le meilleur des cas) dans la zone métastable.

Les paramètres influençant la solubilité des protéines sont très nombreux et variables d'une protéine à l'autre. En pratique, pour trouver un état de sursaturation qui conduise à la cristallisation, il faut tester et maîtriser plusieurs facteurs dont la pureté de l'échantillon protéique, sa dispersité, la composition de la solution tampon dans laquelle est stockée la protéine (pH, sels, co-solvants...), la nature et la concentration des agents précipitants et des additifs ainsi que la température.

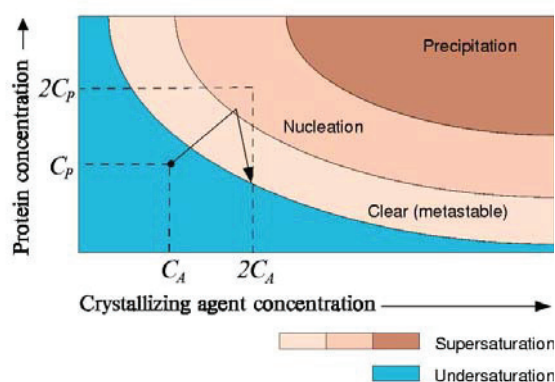


Figure 27 : Diagramme de solubilité suivant la concentration en protéine et la concentration en agent précipitant. La courbe de solubilité sépare la zone où la protéine est en solution (zone de sous-saturation) des zones de sursaturation (zones métastable, de nucléation ou de précipitation) d'après www-structmed.cimr.cam.ac.uk.

4.2.1.B. La diffusion en phase vapeur

Principe : La méthode la plus couramment utilisée pour la cristallogénèse est la diffusion de vapeur (Wlodawer *et al.* 1975). Cette technique permet ainsi de réaliser la transition entre un état de solubilité et un état de sursaturation et est basée sur les propriétés thermodynamiques qui imposent aux concentrations en agents précipitants de deux solutions, enfermées dans un milieu hermétiquement clos, de tendre vers un équilibre par diffusion de vapeur de solvant (**figure 28**). On considère un système clos, constitué d'un réservoir contenant un agent précipitant à la concentration c_1 et d'une goutte contenant ce même agent précipitant à la concentration $c_2 < c_1$ et la protéine à la concentration c' . La diffusion de vapeur à l'intérieur d'un système provoque un transfert net d'eau de la goutte (moins concentrée en agent précipitant) vers le réservoir (plus concentré en agent précipitant) et ce jusqu'à l'obtention d'un état d'équilibre. Ceci a pour conséquence une diminution du volume de la goutte et l'augmentation des concentrations en précipitant et en protéines, aboutissant à une sortie progressive de la protéine de sa zone de solubilité.

Protocole : La technique de la goutte suspendue et de la goutte assise sont utilisées pour la cristallisation de nos protéines d'intérêt. Des essais en gouttes assises ou en gouttes suspendues sont réalisés manuellement dans des boîtes de cristallisation 24 puits (*Greiner Bio-One*) avec des gouttes d'un volume total de 2 μ l minimum. De nombreux essais sont aussi effectués de manière robotisée selon le principe de la goutte assise en boîte 96 puits (*Greiner Bio-One*). Le robot de nanopipetage Mosquito (plateforme cristallogénèse IBCP, *TTP Labtech*) parvient à prélever des volumes compris entre 150 nl et 1,2 μ l. Il permet donc de faire de substantielles économies de matériel biologique et de cribler de nombreuses conditions de cristallisation. Un robot de visualisation RockImager 182 thermostaté à 19 °C (plateforme cristallogénèse IBCP, *Formulatrix*) a également été utilisé.

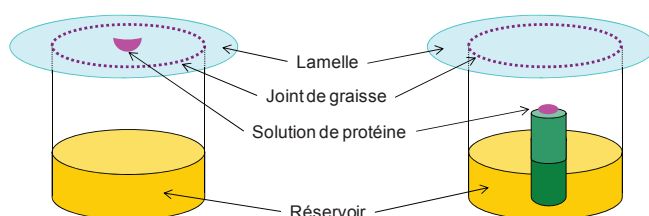


Figure 28 : Croissance cristalline par le principe de la diffusion en phase vapeur, selon les techniques de la goutte suspendue (**gauche**) ou de la goutte assise (**droite**).

4.2.1.C. Techniques de trempage et de co-cristallisation

Principe : Dans le but d'obtenir des cristaux de protéine en complexe avec des substrats, analogues de substrat, des ligands, des ions ou des métaux lourds, deux méthodes peuvent être mises en œuvre : le trempage et la co-cristallisation. Le trempage consiste à introduire ces composés dans le cristal de protéine déjà formé. Ainsi, le ou les produits diffusent dans le cristal par le biais des canaux de solvant présents dans l'édifice cristallin. Si les sites de fixation sont accessibles (ceci dépendant de l'empilement cristallin), la ou les molécules diffusantes peuvent alors les atteindre et éventuellement s'y fixer. La concentration de cette substance et le temps de contact avec le cristal sont déterminés expérimentalement.

Lorsque la méthode de trempage échoue, une technique alternative peut être utilisée : la co-cristallisation. Le ou les produits sont ajoutés à la solution de protéine ou dans la goutte de

cristallisation, préalablement à la formation de cristaux. Les conditions de cristallisation du complexe ainsi obtenu peuvent cependant différer des conditions de cristallisation de la protéine seule (et conduire à des cristaux appartenant à des groupes d'espaces différents) et des étapes d'optimisation peuvent être nécessaires.

4.2.1.D. Cryoprotection des cristaux

Principe : Une fois le cristal exposé au faisceau de rayons X, une série de réactions chimiques comme la formation de radicaux libres et des phénomènes d'ionisation vont pouvoir l'endommager. Ces dégradations causées par le rayonnement X sont importantes si l'on utilise des sources de rayonnement synchrotron. La conduite des expériences à 100 K est une solution très efficace pour résoudre ce problème : le cristal est instantanément gelé dans de l'azote liquide ou gazeux afin de prévenir la formation de glace dans le milieu aqueux. Celui-ci est alors maintenu à une température de 100 K par un flux d'azote gazeux durant toute la phase d'enregistrement des données. Ces conditions cryogéniques impliquent l'utilisation de cryoprotectants qui empêchent la formation de cristaux d'eau au sein du cristal de protéine. Cette cryoprotection peut classiquement s'effectuer en utilisant des concentrations de 5 à 30 % V/V d'éthylène glycol ou de glycérol (d'autres substances pouvant être utilisées) que l'on va laisser rapidement diffuser dans le cristal.

4.2.1.E. Production des rayons X et synchrotron

Principe : Un synchrotron est une source de rayons X bien plus performante qu'un générateur à anode tournante. Il peut être divisé en quatre parties. Le premier élément est un accélérateur linéaire (*Linac*) dans lequel un paquet d'électrons est accéléré jusqu'à une vitesse proche de celle de la lumière. Puis le faisceau d'électrons passe dans un accélérateur circulaire, le *booster*, dans lequel ils sont accélérés jusqu'à atteindre une énergie de 6 GeV à l'installation européenne de rayonnement synchrotron (ESRF) de Grenoble. Une fois que les électrons ont atteint l'énergie voulue, ils sont injectés dans un anneau beaucoup plus grand que le *booster*, l'anneau de stockage, où ils vont faire des centaines de milliers de tours par seconde. L'anneau de stockage est composé de sections courbes constituées par des aimants de courbure et de sections droites correspondant aux portions situées entre deux aimants de courbure. Lorsque les électrons passent dans ces aimants de courbure *Bending Magnet*, ils sont déviés de plusieurs degrés et décrivent ainsi une trajectoire courbe. C'est cette déviation qui conduit les électrons à émettre de la lumière sous forme de photons X. Au niveau des sections droites de cet anneau de stockage, il y a deux autres types d'aimants, les onduleurs *Insertion Device* qui forcent les électrons à suivre une trajectoire ondulée et qui sont à l'origine de la haute brillance du rayonnement synchrotron, et les aimants de focalisation qui concentrent le faisceau d'électrons afin de le rendre le plus fin possible, ce qui permet de produire des rayons X extrêmement brillants. Les faisceaux de rayonnement synchrotron émis par les électrons sont dirigés vers les lignes de lumière, disposées tout autour de l'anneau. Chaque ligne de lumière comporte les instruments d'optique permettant de modifier les caractéristiques de la lumière ainsi que le dispositif nécessaire à l'analyse de l'échantillon et à l'enregistrement des informations générées au moment de l'interaction de la lumière avec le cristal. Selon la localisation de la ligne de lumière, au niveau d'un aimant de courbure ou d'un onduleur, les lignes sont dénommées BM pour *Bending Magnet* ou ID pour *Insertion Device*.

Protocole : Les expériences de diffraction des rayons X se sont déroulées sur plusieurs lignes de lumière (ID et BM) à l'ESRF (*European Synchrotron Radiation Facility*, Grenoble) ainsi qu'au SLS (*Swiss Light Source*, Villigen, Suisse). Un rayonnement synchrotron permet, dans la plupart des cas, d'obtenir de meilleures données qu'avec un générateur à anode tournante. En effet, un faisceau de rayons X issu d'un rayonnement synchrotron est caractérisé par une haute brillance et un diamètre pouvant atteindre seulement quelques dizaines de micromètres. De plus, la collecte des données est nettement plus rapide (près d'une seconde par degré d'oscillation).

4.2.2. Obtention et enregistrement des données de diffraction

4.2.2.A. Principe de diffraction

Principe : Un atome irradié par des rayons X se comporte comme une source secondaire : il diffuse le rayonnement X. Cette diffusion résulte de l'interaction des rayons X avec les électrons des atomes. Si l'atome en question appartient à un cristal, les ondes diffusées interagissent entre elles pour créer des interférences. Les rayons X sont alors émis dans des directions privilégiées : c'est le phénomène de diffraction.

Dans le cas de l'interaction d'un faisceau de rayons X avec un cristal de matière, l'onde diffractée dans une direction h, k, l est représentée par un nombre complexe, $F(hkl)$, appelé facteur de structure et relié au pouvoir de diffusion de l'ensemble des atomes x_j, y_j, z_j , de la maille du cristal. La relation entre le facteur de structure et les atomes diffusants est la suivante où $|F(hkl)|$ et φ_{hkl} sont l'amplitude et la phase du facteur de structure, f_j le facteur de diffusion de l'atome j qui dépend du nombre d'électrons Z_j de chaque atome de la maille :

$$F(hkl) = |F(hkl)| \cdot e^{i\varphi_{hkl}} \quad \text{ou} \quad F(hkl) = \sum_{j=1}^{j=n} f_j \cdot e^{i2\pi(hx_j + ky_j + lz_j)}$$

Si l'on arrive à mesurer expérimentalement les facteurs de structure $F(hkl)$ d'un cristal, il est possible de calculer sa densité électronique (répartition des électrons dans le cristal par unité de volume). En effet la densité électronique $\rho(x,y,z)$ en un point quelconque de coordonnées (x,y,z) de ce cristal, est la transformée de Fourier inverse du facteur de structure :

$$\rho(x, y, z) = \frac{1}{V} \sum_{hkl} F(hkl) \cdot e^{-i2\pi(hx + ky + lz)} \quad \text{où } V \text{ est le volume de la maille du cristal.}$$

Expérimentalement, seules sont enregistrées les intensités $I(hkl)$ des ondes diffractées avec la relation :

$$I(hkl) = |F(hkl)|^2$$

$I(hkl)$, quantité mesurée expérimentalement par un détecteur, est donc proportionnelle au carré du module du facteur de structure $|F(hkl)|^2$. Ainsi, seules les amplitudes des ondes diffusées sont accessibles à l'expérience car les ondes diffractées par un cristal ne peuvent être refocalisées et les phases sont perdues. La détermination de la phase ϕ_{hkl} de chaque réflexion constitue donc l'une des difficultés majeures de la cristallographie.

4.2.2.B. Traitement des données

Principe : Les données de diffraction correspondent à l'ensemble des images enregistrées pour un cristal. Chaque image comporte des tâches de diffraction correspondant à l'imprégnation de photons sur le photodétecteur et dont l'intensité est proportionnelle à la quantité de photons détectée. L'image obtenue à partir de la figure de diffraction correspond au réseau réciproque du cristal. Le traitement des données de diffraction s'effectue en trois étapes. La première est l'indexation des données au cours de laquelle les tâches de diffraction sont localisées, puis des coordonnées leur sont attribuées sous forme de triplet (h,k,l) , où h, k, l sont les indices du réseau réciproque du cristal, caractérisant ainsi une réflexion. Tout ceci se fait de façon informatisée et le logiciel d'indexation détermine les paramètres de maille du cristal et propose ensuite des groupes d'espaces compatibles. La deuxième étape est l'intégration de l'intensité des tâches de diffraction. L'intensité de chaque tâche est mesurée et le bruit de fond au voisinage de chaque tâche est évalué puis retranché à l'intensité. La dernière étape est la mise à l'échelle et la réduction des données. Les intensités mesurées pour une même réflexion peuvent varier d'une image à l'autre, c'est pourquoi ces mesures sont ramenées à une même échelle. Puis les intensités de toutes les réflexions équivalentes, par rapport à la symétrie du cristal, sont moyennées de façon à obtenir un ensemble de réflexions uniques. L'ensemble de ces étapes a été réalisé avec les programmes XDS et XSCALE (Kabsch 2010). La qualité du jeu de données peut être estimée par le facteur d'accord R_{sym} qui permet d'évaluer l'écart entre les intensités des réflexions équivalentes $I_{hkl,j}$ et leur moyenne

$\langle I_{hkl} \rangle$:

$$R_{sym} = \frac{\sum_{hkl} \sum_j |I_{hkl,j} - \langle I_{hkl} \rangle|}{\sum_{hkl} \sum_j I_{hkl,j}}$$

La qualité d'un jeu de données de diffraction est également déterminée par d'autres critères statistiques calculés au cours du traitement. Les paramètres les plus significatifs et utilisés sont :

- la résolution : qui déterminera la finesse des détails dans l'image de la densité électronique
- le signal sur bruit (I/σ) moyen : plus il est grand et meilleure est l'estimation des facteurs de structure
- la complétude : qui mesure le rapport entre le nombre de réflexions uniques mesurées et le nombre théorique de ces réflexions pour la résolution des données. Elle doit être la plus grande possible et supérieure à 90 %
- la redondance : qui correspond au nombre de fois où l'intensité d'une réflexion unique est mesurée. Ce qui est équivalent au nombre moyen de réflexions symétriques mesurées. Plus la valeur numérique de ce critère est grande, meilleure sera l'estimation de l'intensité moyenne d'une réflexion unique en diminuant l'erreur commise à condition que le cristal ne se dégrade pas pendant la collecte.

4.2.3. Le problème des phases

4.2.3.A. La méthode du remplacement isomorphe multiple

Principe : La méthode du remplacement isomorphe multiple utilise le fort pouvoir de diffraction d'atomes liés à la protéine et possédant un nombre élevé d'électrons. Ces atomes sont appelés « atomes lourds ». Ils peuvent être introduits dans la protéine par diffusion à travers le cristal (trempage dans des sels de mercure, de platine, ...) ou par co-cristallisation. Des enregistrements sur au moins un cristal de la protéine native et avec deux cristaux isomorphes comprenant des atomes lourds de type différents, liés sur des sites distincts, sont nécessaires pour déterminer les phases.

4.2.3.B. La méthode de dispersion anormale multiple

Principe : Cette seconde méthode exploite la variation du facteur de diffusion d'un seul type d'atome en fonction de la longueur d'onde des rayons X. En effet, lorsque la longueur d'onde des rayons X correspond à une transition entre deux niveaux d'énergie atomique, il se produit un phénomène de résonance qui se traduit principalement par l'apparition d'un terme complexe dans le facteur de diffusion de l'atome résonant : c'est le phénomène de diffusion anormale.

Si la protéine contient un diffuseur anormal, la différence d'intensité entre les paires de Bijvoet $|F_h(+)|^2$ et $|F_h(-)|^2$ peut être utilisée pour résoudre le problème des phases. En pratique, on utilise la diffusion anormale d'un atome de la protéine (par exemple le sélénium dans le cas de protéines séléno-méthionylées) ou d'un atome introduit dans la protéine par trempage ou co-cristallisation (métal lourd, par exemple) pour obtenir l'information de phases par la mesure des facteurs de structure à au moins trois longueurs d'onde différentes entourant la longueur d'onde d'absorption de l'atome anormal (méthode de dispersion anormale multiple ou MAD). Sous certaines conditions, une seule longueur d'onde peut suffire, c'est la méthode SAD.

4.2.3.C. La méthode du remplacement moléculaire

Principe : Le remplacement moléculaire nécessite l'enregistrement des données de diffraction d'un cristal de la protéine et l'existence d'une structure tridimensionnelle d'une protéine suffisamment homologue. Ainsi, cette structure homologue (appelée modèle-guide) est placée dans la maille de la structure à déterminer et les phases de la macromolécule étudiée sont estimées à partir des facteurs de structure calculés à partir du modèle-guide. Une recherche sur six dimensions est nécessaire afin de trouver le meilleur placement du modèle-guide, à savoir une matrice de rotation et un vecteur de translation. En fait, il est possible de diviser cette recherche en deux étapes à trois dimensions : une recherche de rotation suivie par une recherche de translation.

Pour le choix du modèle final, seuls les opérateurs qui, appliqués au modèle-guide, donneront le meilleur accord entre les facteurs de structure calculés pour le modèle-guide et ceux mesurés pour la structure à déterminer seront sélectionnés.

4.2.3.C.α. Fonction de rotation

Principe : A partir des amplitudes observées (F_{obs}) et de celles calculées (F_{calc}), il est possible d'établir les fonctions de Patterson du modèle-guide et des données de diffraction :

$$P(t) = \frac{1}{V} \sum F(h)^2 \cdot e^{(-i2\pi ht)}$$

avec $F = F_{calc}$ ou F_{obs} , t représente les valeurs interatomiques et h un vecteur dans l'espace réciproque. Les vecteurs interatomiques peuvent être intramoléculaires ou intermoléculaires. Un rayon d'intégration (généralement le rayon du modèle-guide) sera choisi de manière à inclure un maximum de vecteurs intramoléculaires tout en excluant un maximum de vecteurs intermoléculaires.

La fonction de rotation permet de déterminer la matrice de rotation. Elle est calculée sur les vecteurs intramoléculaires par superposition de la Patterson du modèle-guide sur celle des données. Les orientations correspondant aux meilleures corrélations seront sélectionnées. L'algorithme utilisé pour calculer cette fonction a été proposé par Rossmann & Blow (Rossmann & Blow 1962).

4.2.3.C.β. Fonction de translation

Principe : Ce calcul consiste à rechercher par des mouvements de translation, dans une unité asymétrique, les positions correspondant aux meilleures corrélations entre les F_{obs} et les F_{calc} . Les déplacements sont basés sur l'utilisation d'une grille adaptée à la résolution maximale (généralement $\frac{1}{4}$ à $\frac{1}{2}$ de la résolution maximale). Plusieurs méthodes de calcul peuvent être utilisées pour cette recherche.

Le programme MOLREP de la suite CCP4 (Vagin & Teplyakov 2010), contraction de *Molecular Replacement*, correspond à un enchaînement de routines permettant de rechercher automatiquement les solutions de rotation puis de translation présentant la meilleure corrélation.

Un facteur R de confiance représentant l'accord existant entre le modèle (F_{calc}) et les données cristallographiques (F_{obs}) peut être calculé à la fin du processus. Il permet de juger de la vraisemblance du modèle déterminé par le programme. Ce facteur se définit comme suit :

$$R = \frac{\sum_h \|F_{obs}(h) - k F_{calc}(h)\|}{\sum_h |F_{obs}(h)|} \quad \text{avec } k \text{ le facteur d'échelle}$$

4.2.4. Affinement des modèles

Principe : Le modèle de la protéine étudiée (F_{calc}) doit expliquer au maximum les données expérimentales (F_{obs}). Pour obtenir le meilleur accord entre les données expérimentales et calculées, il est nécessaire, pour chaque atome, de modifier sa position spatiale (x,y,z), son facteur d'agitation thermique B et éventuellement son facteur d'occupation. Pour affiner tous ces paramètres, le nombre d'observations doit normalement être supérieur au nombre de paramètres à affiner d'au moins un facteur trois à quatre, du fait des erreurs sur les mesures. Ce nombre d'observations est rarement atteint en cristallographie des macromolécules biologiques (résolution insuffisante) et l'affinement utilise des restrictions stéréochimiques en plus des contraintes cristallographiques pour minimiser l'énergie totale du système selon l'équation :

$$E_{totale} = E_{cristallographique} + [E_{bond} + E_{angl} + E_{impr} + E_{elec} + E_{vdw} + E_{pelec} + E_{pvdw}]$$

ou de manière simplifiée :

$$E_{totale} = E_{cristallographique} + E_{empirique}$$

Le terme $E_{empirique}$ représente les termes d'énergie liés aux contraintes stéréochimiques. Il est égal à la somme des énergies liées aux interactions covalentes (longueur E_{bond} , angles de liaison E_{angl} , angles impropres E_{impr}) et non covalentes (potentiels électrostatiques E_{elec} et interactions de Van der Waals E_{vdw}). Pour les atomes reliés par symétrie cristalline, ces interactions non covalentes sont représentées par les termes E_{pelec} et E_{pvdw} . L'évolution de l'énergie empirique est suivie en considérant la différence entre la valeur actuelle et la valeur idéale. Ainsi, si les atomes dévient de la géométrie idéale, la quantité d'énergie augmente.

Le terme $E_{cristallographique}$ est la différence pondérée existant entre les facteurs de structure observés et calculés. Il s'écrit :

$$E_{cristallographique} = \frac{W_a}{N_a} \sum_{hkl} W_{hkl} [|F_{obs}(hkl)| - k \cdot |F_{calc}(hkl)|]$$

Le facteur W_a est un facteur de pondération appliqué à toutes les réflexions de manière à ce que le gradient de l'énergie soit égal au gradient de l'énergie empirique. Il est déterminé par une étape de dynamique moléculaire. Le facteur N_a est un facteur de normalisation, W_{hkl} le poids associé à une réflexion donnée, et k est le facteur de mise à l'échelle entre les facteurs de structure observés et calculés.

Les cycles d'affinement permettent, par de subtiles variations de coordonnées atomiques, de faire évoluer le système vers une conformation plus stable et donc de moindre énergie.

Pour réaliser les cycles d'affinement, nous avons utilisé le logiciel REFMAC (Collaborative Computational Project 1994) ou PHENIX (Adams *et al.* 2010). Ces programmes permettent d'appliquer des contraintes stéréochimiques au modèle, de manière à maintenir la longueur des liaisons, les angles de valence, etc. proches des valeurs canoniques. Aux résolutions usuelles pour les structures de protéines (3,0 à 1,5 Å), le nombre d'observations est insuffisant pour effectuer un

affinement sans contrainte. Tout au long de l'affinement, la qualité du modèle est contrôlée par l'évolution des facteurs R et R_{free} (définis ci-après) mais aussi par la valeur des écarts types par rapport aux distances stéréochimiques idéales.

$$R = \frac{\sum_h \|F_{obs}(h) - k|F_{calc}(h)\|}{\sum_h |F_{obs}(h)|} \quad \text{et} \quad R_{free} = \frac{\sum_{h \in T} \|F_{obs}(h) - k|F_{calc}(h)\|}{\sum_{h \in T} |F_{obs}(h)|}$$

où k est un facteur d'échelle et le R_{free} est calculé par le rapport à un jeu de 5 à 10 % de réflexions qui n'interviennent pas dans l'affinement (Brünger 1992).

Le facteur R_{free} se calcule comme le facteur R mais porte sur des réflexions non biaisées par le modèle, dans la mesure où elles ne sont pas prises en compte dans l'affinement. Un modèle correct doit pouvoir rester en accord avec ces réflexions tout au long de l'affinement. Pour l'affinement de nos structures, 5 % des réflexions ont été aléatoirement mises de côté pour ce calcul.

4.2.5. Les cartes de densité électronique

Principe : La seule application des méthodes d'affinement ne suffit pas à obtenir une structure affinée et un certain nombre de corrections doivent être faites manuellement sur le modèle. De plus, ce dernier est souvent partiel en début d'affinement et doit être complété. Les modifications et les ajouts sont effectués à partir de l'interprétation de différents types de cartes de densité électronique, notamment des cartes dites de « Fourier difference ». Celles-ci sont calculées par transformée de Fourier de différentes combinaisons entre facteurs de structures observés et ceux calculés à partir du modèle :

$$\rho(x, y, z) = \frac{1}{V} \sum_h \sum_k \sum_l (m|F_{obs}| - n|F_{calc}|) e^{i\varphi_{calc}} e^{-2\pi i(hx + ky + lz)}$$

où φ_{calc} représente la phase calculée et (m, n) appartient à {(3, 2), (2, 1), (1, 1)}.

Les cartes calculées avec le couple (3, 2) et (2, 1) sont en général utilisées pour visualiser la densité électronique de la structure et construire le modèle car elles donnent plus de poids aux observations expérimentales et réduisent le biais du modèle. Les différentes cartes ont été observées avec le programme COOT (Emsley & Cowtan 2004). Les cartes de « Fourier difference », c'est-à-dire celles calculées avec les coefficients (1, 1) permettent de localiser les différences entre le modèle et la structure. Les pics positifs de cette carte indiquent la position d'atomes manquants dans le modèle. A l'inverse, les pics négatifs indiquent les positions d'atomes excédentaires. Une paire de pics positif-négatif signale qu'un atome, présent dans le modèle, est mal placé et qu'il doit être déplacé du pic négatif vers le pic positif. Cependant, dans un cas réel, l'interprétation de ces cartes est plus complexe car un certain nombre d'erreurs ou d'artéfacts peuvent entacher la carte. On parle alors de « bruit ».

RÉSULTATS ET DISCUSSION

1. Intégrase IN-CCD-H103C

1.1. Modélisation moléculaire

Les premières études réalisées au sein des équipes de BioCristallographie de l'IBCP et de Rétrovirologie de l'INRA sur le site de Lyon-Gerland, ont mis en évidence que le domaine catalytique sauvage de l'intégrase du RAV-1 (IN-CCD-WT) était monomérique en solution même en présence de MES, et que l'ajout de zinc aboutissait à la précipitation de la protéine (manuscrit de thèse d'A. Ballandras, 2010). Une stratégie a donc été mise en place afin de démontrer l'existence de la « nouvelle » structure quaternaire en solution et indépendamment de la présence de MES ou de zinc.

L'observation de la structure cristalline d'IN-CCD-WT sous forme dimérique a mis en évidence que ce dimère était stabilisé par les histidines H103A et H103B de chaque monomère, qui coordonnent un ion zinc au niveau de l'interface (**figure 29, gauche**). Nous avons donc eu l'idée de muter ces deux histidines en cystéines et une modélisation moléculaire a permis de mettre en évidence que les deux cystéines de chaque monomère seraient assez proches pour permettre la formation d'un pont disulfure inter-moléculaire sur le « nouvel » assemblage (**figure 29, droite**).

Nous avons donc initié la surproduction du mutant H103C (IN-CCD-H103C) pour tenter de visualiser la formation d'une espèce dimérique stabilisée par un pont disulfure C103-C103.

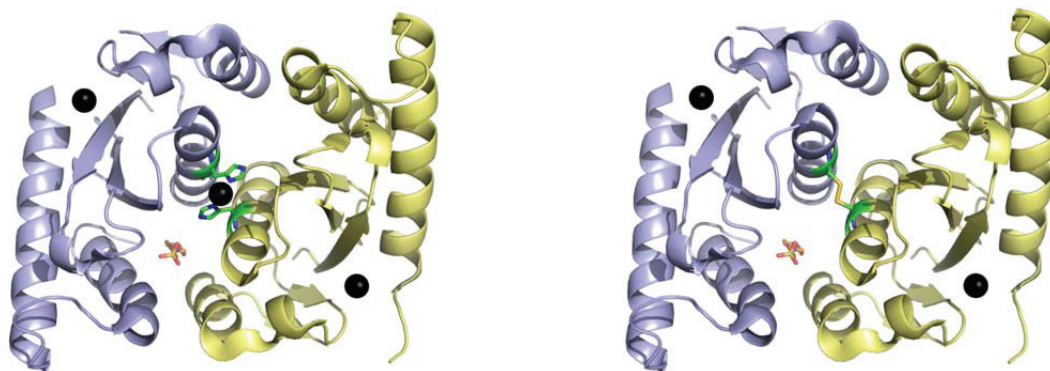


Figure 29 : (**Gauche**) Structure cristalline d'IN-CCD-WT sous forme dimérique avec la nouvelle interface à trois paires d'hélices α . Les histidines 103 sont en vert, les ions zinc en noir et la molécule de MES en orange.

(**Droite**) Modèle d'IN-CCD-H103C sous forme dimérique avec la nouvelle interface et le pont disulfure modélisé entre les cystéines 103. Les ions zinc et le MES sont également représentés.

1.2. Surproduction et purification

Afin de vérifier la formation d'un pont disulfure inter-moléculaire entre les cystéines 103 du mutant IN-CCD-H103C, une stratégie de surproduction en système procaryote (*E. coli*), de solubilisation et de purification de la protéine mutée a été mise en place.

La surproduction d'IN-CCD-H103C a été réalisée dans des bactéries Rosetta-gami 2 DE3 pLysS transformées par le plasmide pETG-10a contenant le gène codant IN-CCD-H103C (obtenu par mutagenèse dirigée par PCR à partir du gène codant IN-CCD-WT). Nous avons décidé d'utiliser ces bactéries, qui sont adaptées à la formation de ponts disulfures dans le cytoplasme, ceci afin de favoriser la formation éventuelle de dimères par la création d'un pont disulfure inter-moléculaire au cours de la surproduction.

Le protocole de surproduction et de solubilisation avait déjà été mis au point au sein du laboratoire pour IN-CCD-WT dans les cellules BL21 DE3 pLysS ; ce même protocole a été employé pour IN-CCD-H103C. Un milieu de culture LB a été utilisé et l'induction de la surproduction a été déclenchée par l'ajout de 1 mM d'IPTG lorsque les bactéries se trouvaient en phase exponentielle de croissance (densité optique à 600 nm comprise entre 0,4 et 0,6). La culture a ensuite été incubée pendant une nuit à 25 °C avant de récupérer les bactéries. La lyse des cellules a été réalisée par sonication dans une solution tampon T1 (20 mM Tris HCl pH 8,0, 500 mM NaCl, 20 mM imidazole). La protéine surproduite, qui possède une étiquette de six histidines, a ensuite été purifiée par chromatographie d'affinité pour le nickel à l'aide d'une colonne possédant une phase solide en résine avec un bras espaceur IDA et chélatée avec du nickel (colonne Ni-IDA). L'absorbance à 280 nm a été suivie tout le long du processus de purification ainsi que la conductance propre à chaque tampon (**figure 30, gauche**).

Le protocole de purification avait également déjà été mis au point pour la forme sauvage et a donc été suivi en portant une attention particulière à omettre tout agent réducteur dans les solutions tampons, ceci afin d'éviter la rupture des éventuels ponts disulfures formés. Après injection de la solution de surnageant sur la colonne et passage de la totalité de la fraction non retenue, un premier lavage dans le tampon T1, suivi d'un lavage dans un tampon T2 de plus forte force ionique (identique au tampon T1 mais contenant 1 M de NaCl), a permis de décrocher de nombreux contaminants fixés de façon aspécifique sur la colonne (protéines et acides nucléiques) (**figure 30, gauche, fractions A et B**). L'élution de la protéine a ensuite été réalisée grâce à un gradient d'imidazole par mélange du tampon T1 avec un tampon T3 (20 mM Tris HCl pH 8,0, 500 mM NaCl, et 1 M imidazole). Notre protéine semble se décrocher pour un gradient d'environ 90 % V/V tampon T1/10 % V/V tampon T3, ce qui correspond à une concentration de près de 100 mM d'imidazole (**figure 30, gauche, fraction C**). L'élution a ensuite été prolongée en maintenant le gradient en palier, ceci afin de limiter la concentration d'imidazole dans les fractions d'élution récupérées.

Après analyse des fractions d'élution par électrophorèse sur gel discontinu de polyacrylamide 12 % en conditions dénaturantes et réductrices (SDS-PAGE), il apparaît qu'IN-CCD-H103C est présent dans les dernières fractions d'élution avec très peu de contaminants (**figure 30, droite**). L'apparition d'une bande majoritaire migrant entre les marqueurs de masses moléculaires de 17 et 26 kDa, est cohérente avec la masse moléculaire attendue d'un monomère d'IN-CCD-H103C calculée à partir de sa séquence protéique (17,5 kDa). Les fractions d'élution contenant la protéine avec le plus haut niveau de pureté ont ensuite été rassemblées puis dialysées afin d'enlever l'imidazole.

La protéine surproduite s'est avérée être présente majoritairement dans la fraction insoluble après la lyse des bactéries et moins de 0,5 mg de protéine par litre de culture a été obtenu en fin de

purification. Ceci constitue un rendement de purification très faible, le rendement obtenu pour IN-CCD-WT étant de près de 4 mg de protéine par litre de culture (manuscrit de thèse d'A. Ballandras, 2010). Du fait de la faible quantité de protéine récupérée, nous avons continué les études préliminaires de caractérisation d'IN-CCD-H103C en solution en conservant l'étiquette de purification.

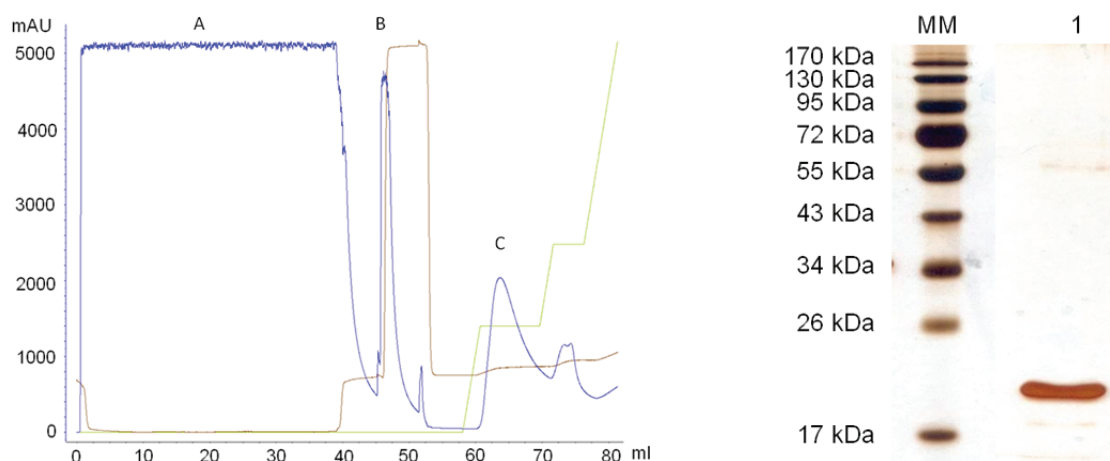


Figure 30 : (Gauche) Chromatogramme de la purification d'IN-CCD-H103C par chromatographie d'affinité pour le nickel. Courbe bleue : suivie de l'absorbance en sortie de colonne pour une longueur d'onde de 280 nm (axe des ordonnées) et en fonction du volume d'élution (axe des abscisses). Courbe marron : suivi de la conductance du tampon en sortie de colonne. Courbe verte : suivi du gradient tampon T1/tampon T3 exprimé en pourcentage V/V de tampon T3. Fraction A : fraction non retenue. Fraction B : fraction de lavage. Fraction C : fraction d'élution.

(Droite) Analyse par électrophorèse sur gel discontinu de polyacrylamide (12 %) en conditions dénaturantes et réductrices (SDS-PAGE) et avec révélation au nitrate d'argent, d'IN-CCD-H103C en fin de purification. Piste MM : marqueurs de masses moléculaires connues. Piste 1 : IN-CCD-H103C purifié.

1.3. Caractérisation de la protéine en solution

IN-CCD-H103C a été précédemment purifié avec un bon niveau de pureté mais en très faible quantité. Nous avons malgré tout poursuivi notre étude de caractérisation de la protéine en solution, et avec son étiquette de purification, afin de déterminer son état oligomérique en fin de purification. En effet, des expériences précédentes de chromatographie d'exclusion de taille, réalisées par le Dr A. Ballandras au cours de sa thèse, avaient montré qu'IN-CCD-WT était majoritairement sous forme monomérique en solution. Nous avons donc voulu vérifier si IN-CCD-H103C était également monomérique ou si la forme dimérique avait pu se former grâce à un pont disulfure inter-moléculaire entre les cystéines 103. Pour cela, l'échantillon d'IN-CCD-H103C obtenu en fin de purification a été analysé par chromatographie d'exclusion de taille (Superdex 75 HR 10/30).

Nous pouvons distinguer sur le profil d'élution obtenu un pic majoritaire, suivi d'un pic de plus faible intensité (figure 31, gauche, pics A et B). La protéine semble donc présente dans ces conditions expérimentales sous deux états oligomériques avec un état oligomérique majoritaire. La nature de ces assemblages a été estimée à partir des volumes d'élution correspondant au maximum du pic et après calibration de la colonne avec des protéines globulaires de masses moléculaires connues. Il apparaît que le volume d'élution correspondant au maximum du pic majoritaire est de près de 11 ml. Ce volume d'élution est très proche de celui obtenu pour la chymotrypsine (11,52 ml) dont la masse moléculaire est de 25 kDa. La droite étalon représentant le logarithme des masses

moléculaires des standards en fonction de leurs volumes d'élution, a permis d'estimer une masse moléculaire proche de 32 kDa. Le volume d'élution correspondant au maximum du pic minoritaire est de près de 13 ml. Ce volume d'élution est très proche de celui obtenu pour la RnaseA (12,9 ml) dont la masse moléculaire est de 13,7 kDa et la droite étalon précédente a permis d'estimer une masse moléculaire proche de 13 kDa. Ces valeurs de masses moléculaires obtenues pour le pic majoritaire et le pic minoritaire sont respectivement proches de la masse moléculaire d'un dimère (35 kDa) et d'un monomère (17,5 kDa) du domaine catalytique de l'intégrase. En sortie de purification, le domaine IN-CCD-H103C serait donc majoritairement sous forme dimérique avec une faible proportion de monomères.

Nous avons ensuite souhaité vérifier par électrophorèse sur gel discontinu de polyacrylamide 12 % en conditions dénaturantes et en absence d'agent réducteur (absence de β -mercaptoéthanol normalement contenu dans le tampon Laemmli), si le dimère était le résultat de la formation d'un pont disulfure inter-moléculaire (**figure 31, droite**). Les fractions correspondant au pic majoritaire et au pic minoritaire ont été incubées avec un tampon Laemmli dépourvu d'agent réducteur et analysées par SDS-PAGE. L'apparition d'une bande majoritaire migrant entre les standards de masses moléculaires de 34 et 43 kDa est cohérente avec la masse moléculaire attendue d'un dimère d'IN-CCD-H103C (35 kDa) (**figure 31, droite, pistes 1 à 3**). Sa visualisation sur le gel réalisé en conditions dénaturantes et non réductrices indique la formation d'un dimère covalent impliquant la formation d'un pont disulfure. Ce résultat est confirmé par l'apparition d'une forme monomérique en présence d'agent réducteur (**figure 30, droite**). De façon intéressante, la fraction correspondant au pic minoritaire indique également la présence d'une forme dimérique (**figure 31, droite, pistes 4 à 5**). Une explication envisageable serait que la totalité de la protéine monomérique se soit progressivement dimérisée entre la fin de la chromatographie d'exclusion et la réalisation de l'analyse SDS-PAGE sous l'effet d'une oxydation par l'oxygène de l'air. Il apparaît donc, à la vue de ces résultats, qu'IN-CCD-H103C surproduit dans la souche bactérienne Rosetta-gami 2 DE3 pLysS est déjà majoritairement sous forme dimérique en fin de purification et que cette forme est bien stabilisée par un pont disulfure formé au cours de la surproduction.

Une autre cystéine (C125) est présente dans le domaine catalytique et nous ne pouvions pas exclure la possibilité d'un pont disulfure différent de celui attendu entre les deux cystéines 103. Des analyses par spectrométrie de masse réalisées en conditions réductrices et non réductrices ont permis de confirmer la présence d'un dipeptide contenant un pont disulfure entre les cystéines C103 (données non montrées).

Ces premiers résultats ont donc mis en évidence que la nouvelle interface pouvait bien se former en solution, indépendamment de la présence de MES ou de zinc, et que cette dernière pouvait être stabilisée par la création d'un pont disulfure. Toutefois, le rendement de purification d'IN-CCD-H103C s'est avéré être très faible ne permettant pas de poursuivre les études biochimiques et les tests d'interactions avec les petites molécules. Par ailleurs, il est apparu que la protéine se dégradait très rapidement pendant son stockage (même à 4 °C). Nous avons donc décidé de modifier le protocole de surproduction (changement de la souche bactérienne pour des BL21 DE3 pLysS), en espérant augmenter le rendement de purification et la stabilité de la protéine.

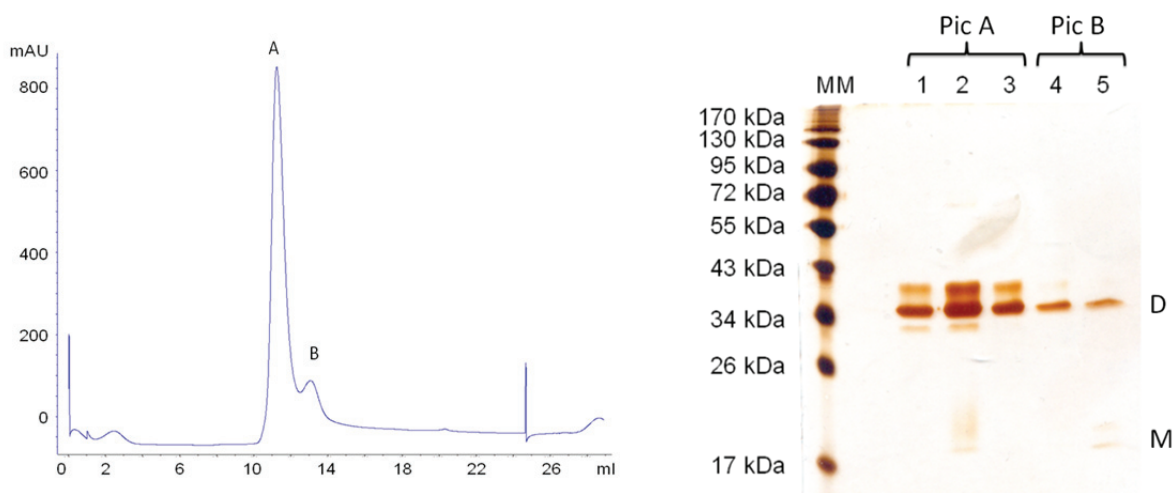


Figure 31 : (**Gauche**) Profil d'élution d'IN-CCD-H103C obtenu par chromatographie d'exclusion de taille (Superdex 75 HR 10/30). Courbe bleue : suivie de l'absorbance en sortie de colonne pour une longueur d'onde de 280 nm (axe des ordonnées) en fonction du volume d'élution (axe des abscisses). (**Droite**) Analyse par électrophorèse sur gel discontinu de polyacrylamide (12 %) en conditions dénaturantes (SDS-PAGE) et en absence d'agent réducteur, avec révélation au nitrate d'argent, des fractions issues de la chromatographie d'exclusion. Piste MM : marqueurs de masses moléculaires connues. Piste 1, 2 et 3 : pic A sur le chromatogramme. Pistes 4 et 5 : pic B sur le chromatogramme. M et D indiquent respectivement le front de migration de la forme monomérique et dimérique.

1.4. Optimisation de la surproduction

Après avoir transformé les cellules BL21 DE3 pLysS par notre plasmide d'intérêt codant IN-CCD-H103C, une surproduction suivie d'une purification ont été réalisées suivant le même protocole que précédemment (**figure 32, gauche**). Il est apparu qu'IN-CCD-H103C était présent dans les fractions d'élution avec de nombreux contaminants (données non montrées). Les fractions d'élution contenant la protéine ont été rassemblées puis dialysées afin d'enlever l'imidazole, avant de réaliser un clivage de l'étiquette de purification par la thrombine (1 unité de thrombine pour 5 mg de protéine, sur la nuit et à 4 °C). Une nouvelle chromatographie d'affinité pour le nickel a ensuite été réalisée et a permis d'éliminer les contaminants qui sont restés accrochés à la colonne ainsi que les étiquettes de purifications clivées, permettant l'obtention d'une protéine avec un haut niveau de pureté.

Malgré une quantité toujours importante de protéine dans la fraction insoluble suite à l'étape de lyse, IN-CCD-H103C semble être obtenu en plus grande quantité par l'utilisation de cette souche bactérienne et près de 3 mg de protéine par litre de culture a été obtenu en fin de purification et après clivage de l'étiquette.

Nous avons ensuite voulu vérifier par électrophorèse sur gel discontinu de polyacrylamide 12 % en conditions dénaturantes et en présence ou absence d'agent réducteur dans le tampon Laemmli, si le dimère précédemment observé et stabilisé par un pont disulfure inter-moléculaire était toujours présent en sortie de purification par l'utilisation de la souche BL21 DE3 pLysS (**figure 32, droite**). L'apparition d'une bande majoritaire en présence et en absence d'agent réducteur et migrant entre les standards de masses moléculaires de 10 et 17 kDa est cohérente avec la masse moléculaire attendue d'un monomère d'IN-CCD-H103C sans étiquette de purification (16,3 kDa) (**figure 32, droite, piste 1**). Une bande minoritaire migrant entre les standards de masses moléculaires de 26 et 34 kDa en absence d'agent réducteur, semble indiquer la présence de la forme dimérique (32,6 kDa)

en très faible proportion (**figure 32, droite, piste 1**). Ce résultat est confirmé par la disparition de cette bande en présence d'agent réducteur (**figure 32, droite, piste 2**).

Il apparaît donc à la vue de ces résultats qu'IN-CCD-H103C surproduit dans la souche bactérienne BL21 DE3 pLysS est majoritairement sous forme monomérique en sortie de purification. Le milieu réducteur des bactéries BL21 DE3 pLysS n'est pas adapté à la formation de ponts disulfures dans le cytoplasme et ceci pourrait expliquer le résultat obtenu. La protéine a cependant été surproduite en plus grande quantité et ne se dégrade pas rapidement pendant son stockage à 4 °C. Nous avons donc décidé de conserver ce protocole et nous avons alors recherché un moyen de former le pont disulfure inter-moléculaire *in vitro* afin de favoriser la formation de l'espèce dimérique.

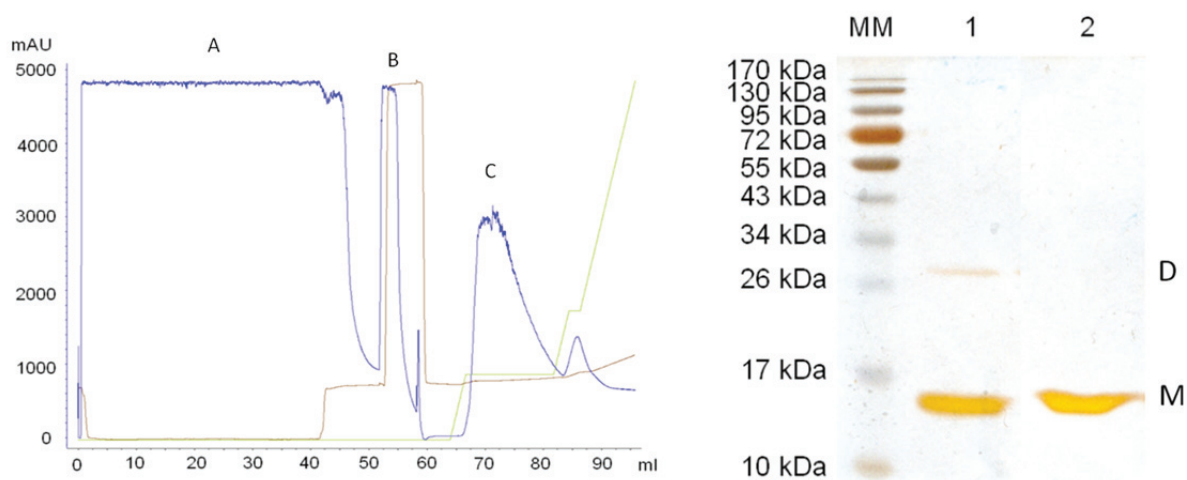


Figure 32 : (**Gauche**) Chromatogramme de la purification d'IN-CCD-H103C par chromatographie d'affinité pour le nickel. Courbe bleue : suivie de l'absorbance en sortie de colonne pour une longueur d'onde de 280 nm (axe des ordonnées) et en fonction du volume d'élution (axe des abscisses). Courbe marron : suivi de la conductance du tampon en sortie de colonne. Courbe verte : suivi du gradient tampon T1/tampon T3 exprimé en pourcentage V/V de tampon T3. Fraction A : fraction non retenue. Fraction B : fraction de lavage. Fraction C : fraction d'élution.

(**Droite**) Analyse par électrophorèse sur gel discontinu de polyacrylamide (12 %) en conditions dénaturantes (SDS-PAGE) et en présence ou absence d'agent réducteur, avec révélation au nitrate d'argent, d'IN-CCD-H103C en fin de purification. Piste MM : marqueurs de masses moléculaires connues. Piste 1 : IN-CCD-H103C purifié et analysé en conditions non réductrices. Piste 2 : IN-CCD-H103C purifié et analysé en conditions réductrices. M et D indiquent respectivement le front de migration de la forme monomérique et dimérique.

1.5. Formation du pont disulfure *in vitro*

Après avoir obtenu IN-CCD-H103C avec un bon rendement de purification, mais sous une forme majoritairement monomérique en solution, nous avons recherché un moyen pour favoriser l'apparition de l'espèce dimérique après les étapes de purification. Une analyse de la littérature a permis de mettre en évidence l'utilisation du peroxyde d'hydrogène (H_2O_2) en tant qu'oxydant pour catalyser la formation de ponts disulfures entre deux cystéines vicinales par la formation d'intermédiaires d'acides sulféniques très réactifs (Paoli *et al.* 2001). Par ailleurs, le pH de la solution tampon d'IN-CCD-H103C étant légèrement basique (8,0), ce dernier peut être bien adapté à la formation de ponts disulfures, les groupements sulfhydryles étant plus réactifs lorsqu'ils sont ionisés (le pKa d'ionisation de la fonction thiol est généralement proche de 8,0).

Des premiers essais ont donc été réalisés avec IN-CCD-H103C purifié (environ 0,2 mg/ml) incubé en présence de peroxyde d'hydrogène (à température ambiante) à différentes concentrations (peroxyde d'hydrogène préparé à 10 mM initial à partir d'un stock à 10 M et utilisé à 0,2-2 mM final) et pendant différents temps (cinq minutes à une heure). Nous avons ensuite contrôlé l'apparition d'une espèce dimérique stabilisée par un pont disulfure par électrophorèse sur gel discontinu de polyacrylamide 12 % en conditions dénaturantes et en présence ou non d'agent réducteur dans le tampon Laemmli (**figure 33**).

La protéine analysée par SDS-PAGE en fin de purification et en absence d'agent réducteur, est caractérisée par une bande majoritaire migrant entre les standards de masses moléculaires de 10 et 17 kDa (**figure 33, piste 1**). Ce résultat est cohérent avec la masse moléculaire attendue d'un monomère du domaine catalytique de l'intégrase sans étiquette de purification, témoignant que IN-CCD-H103C est bien majoritairement sous forme monomérique suite à la purification. Il apparaît cependant qu'en présence de peroxyde d'hydrogène à 0,2 mM, une bande migrant entre les standards de masses moléculaires de 26 et 34 kDa et correspondant à la forme dimérique, apparaît sur le gel après cinq minutes d'incubation et devient majoritaire devant la forme monomérique au bout de trente minutes ou une heure d'incubation (**figure 33, pistes 2 à 4**). Afin de vérifier si la concentration en peroxyde d'hydrogène pouvait être limitante pour la réaction de formation du pont disulfure, la protéine a également été incubée avec deux concentrations plus fortes de peroxyde d'hydrogène (1 mM et 2 mM) pendant trente minutes. Il apparaît sur l'analyse SDS-PAGE que la concentration en peroxyde d'hydrogène peut être limitante et que la réaction de formation du pont disulfure est optimale pour une concentration de 1 mM de peroxyde d'hydrogène et pour un temps d'incubation d'au moins trente minutes (**figure 33, pistes 5 à 7**).

Dans le but de vérifier que l'apparition de la forme dimérique n'était pas le résultat d'un artéfact lié à l'oxydation de la protéine et à la formation de liaisons covalentes entre des résidus oxydés, la réversibilité de cet assemblage a été évaluée par une analyse SDS-PAGE en présence d'agent réducteur. L'incubation de la protéine préalablement traitée par le peroxyde d'hydrogène avec un agent réducteur est caractérisée sur gel par l'apparition d'une seule bande correspondant à la masse moléculaire attendue d'un monomère (**figure 33, pistes 8 et 9**).

Il apparaît donc bien à la vue de ces résultats que le peroxyde d'hydrogène a permis de favoriser l'apparition du dimère stabilisé par un pont disulfure réversible par l'ajout d'un agent réducteur. La confirmation de l'implication des cystéines 103 dans ce pont disulfure a été faite par spectrométrie de masse. Par ailleurs, le traitement par le peroxyde d'hydrogène n'a pas permis l'apparition d'une forme dimérique sur IN-CCD-WT (données non montrées).

Afin de vérifier l'intégrité de la protéine après ce traitement, l'activité de désintégration d'IN-CCD-H103C traité avec le peroxyde d'hydrogène a été testée par le Dr C. Ronfort. Il est apparu que IN-CCD-H103C était toujours bien actif pour l'activité de désintégration après ce traitement (données non montrées).

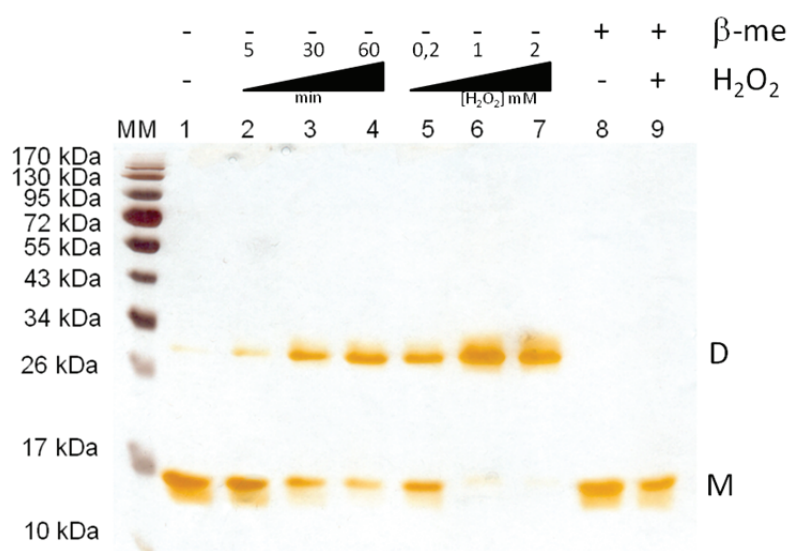


Figure 33 : Analyse par électrophorèse sur gel discontinu de polyacrylamide (12 %) en conditions dénaturantes et en présence ou absence d'agent réducteur d'IN-CCD-H103C purifié, avec révélation au nitrate d'argent. Piste MM : marqueurs de masses moléculaires connues. Piste 1 : IN-CCD-H103C purifié et analysé en conditions non réductrices (et sans peroxyde d'hydrogène). Pistes 2, 3 et 4 : IN-CCD-H103C incubé respectivement cinq, trente et soixante minutes avec 0,2 mM H₂O₂ et analysé en conditions non réductrices. Pistes 5, 6 et 7 : IN-CCD-H103C incubé respectivement avec 0,2, 1 et 2 mM H₂O₂ pendant trente minutes et analysé en conditions non réductrices. Piste 8 et 9 : IN-CCD-H103C respectivement non incubé et incubé trente minutes avec 0,2 mM H₂O₂ et analysé en conditions réductrices. M et D indiquent respectivement le front de migration de la forme monomérique et dimérique.

1.6. Essais de cristallisation

Nous avons tenté d'obtenir des cristaux de la forme dimérique d'IN-CCD-H103C stabilisée par un pont disulfure, par l'utilisation des conditions de cristallisation déjà définies pour IN-CCD-WT (manuscrit de thèse d'A. Ballandras, 2010). Après plusieurs jours, des cristaux sont apparus dans les mêmes conditions contenant le MES et le zinc (10 mM ZnCl₂, 20 % w/v PEG 6 000, 100 mM MES, pH 6,0, 16 °C) (**figure 34**).

Malheureusement, ces cristaux n'ont pu être testés sur un rayonnement synchrotron, ces derniers ayant disparus au bout de plusieurs jours. De nouveaux essais de cristallisation ont été initiés dans ces mêmes conditions, mais aucun nouveau cristal de protéine n'a pu être observé à ce stade de l'étude. Il est possible que la stabilisation préalable de la forme dimérique et la mutation de l'histidine 103 en cystéine ait affecté les conditions de cristallisation. La cristallisation de l'assemblage dimérique stabilisé pourrait ainsi nécessiter un criblage de nouvelles conditions de cristallisation.

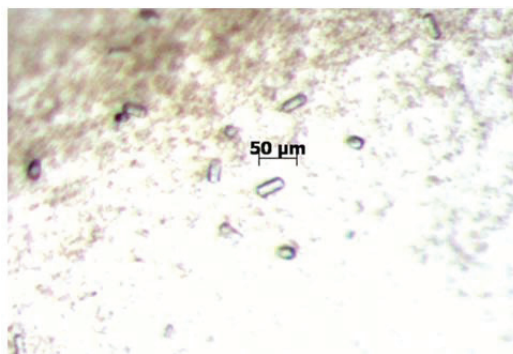


Figure 34 : Cristaux obtenus pour IN-CCD-H103C dans la condition 10 mM ZnCl_2 , 20 % w/v PEG 6 000, 100 mM MES, pH 6,0, 16 °C.

1.7. Interactions de petites molécules avec le nouvel assemblage

1.7.1. Conception des molécules

Les études préliminaires de la cavité située dans la nouvelle interface dimérique d'IN-CCD-WT nous ont permis de déterminer ses différentes propriétés physico-chimiques telles que sa taille, forme, ainsi que les types d'interactions possibles (interactions polaires/non polaires, contacts hydrophobes, etc) et les surfaces accessibles au solvant dont les charges électrostatiques résiduelles. Nous avons ainsi pu déterminer que la moitié de la cavité était plutôt chargée positivement tandis que l'autre moitié était plutôt non chargée. De nombreux résidus polaires semblent par ailleurs être en mesure d'établir des liaisons dipôle/dipôle ainsi que des liaisons hydrogènes avec un ligand. Enfin, il apparaît que la cavité a une forme légèrement coudée et fait environ 18 Å de long, là où le MES n'en utilise que 8.

Dans un premier temps, nous avons orienté notre recherche sur une première série de molécules directement dérivées de l'acide 2-(4-Morpholino)-Ethane Sufonique (MES), puisque cette molécule était observée dans la crevasse de la nouvelle interface dimérique. Nos collaborateurs chimistes ont ainsi synthétisé une quinzaine de molécules analogues au MES et présentant des variations soit sur la nature de l'amine (amines cycliques, acycliques, aromatiques, benzyliques), soit sur la nature du motif anionique (analogues de type carboxylate ou tétrazoly), soit sur la longueur de la chaîne carbonée (**figure 35**).

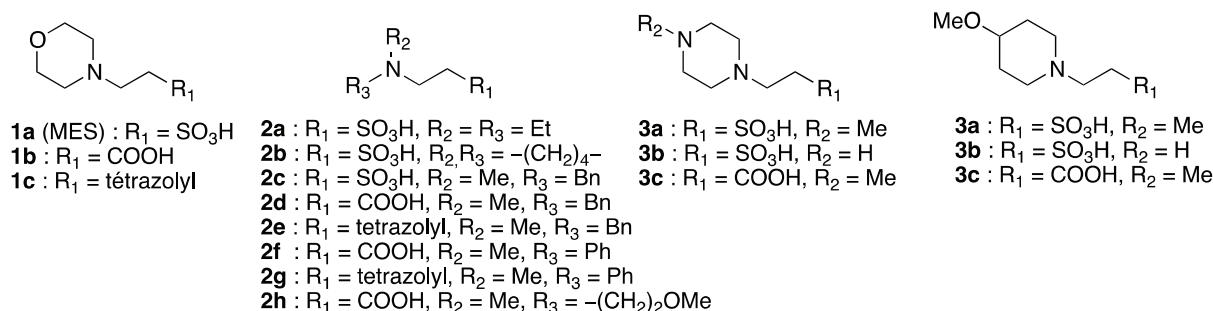


Figure 35 : Molécules analogues au MES et synthétisées par l'équipe du Dr Philippe Maurin.

Suite à cette synthèse, les effets de ces composés ont été évalués *in vitro* sur les activités catalytiques de l'intégrase (activités de clivage et de transfert de brins) et *in cellulo* sur la réplication du virus RAV-1. L'équipe du Dr Corinne Ronfort a pu montrer que ces composés n'avaient pas d'effet inhibiteur à une concentration maximale testée de 100 μM et en présence de zinc (10 μM), sur les activités de clivage et de transfert de brins de l'intégrase du RAV-1. Par ailleurs, ces composés n'ont pas montré d'effet inhibiteur sur la réplication du virus RAV-1 (sur des cellules fibroblastiques de caille ou cellules QT6), à une dose de toxicité de 20 % (concentrations de composés comprises entre 1 et 25 mM) et pendant un intervalle de 24 à 72 H. Les développements sur ces composés n'ont donc pas été poursuivis.

Au cours de ces essais, il nous est apparu que la cavité serait capable d'interagir avec des molécules de plus grande taille que le MES (deux à trois fois supérieure), plutôt polaires, avec de nombreux hétéroatomes pouvant être des donneurs ou accepteurs de liaisons hydrogènes. Pendant l'année 2012, un étudiant du Master de Chimie de l'ENS de Lyon, Nils Aronsohn, a effectué son stage de recherche (encadré par le Pr. Patrice Gouet et le Dr Philippe Maurin du laboratoire de Chimie de l'ENS) sur la synthèse d'inhibiteurs allostériques potentiels ciblant la nouvelle interface du CCD d'IN-RAV-1. Cet étudiant a observé que le Raltégravir pourrait présenter les caractéristiques décrites ci-dessus et confirmé par des expériences *in silico* (*docking*, programme AutoDock/Vina) que ce dernier possédait un très bon score de fixation et occupait pratiquement toute la cavité (**figure 36, haut, gauche**). Les résultats de *docking* du Raltégravir prévoient une énergie libre de liaison de $-8,4 \text{ Kcal.mol}^{-1}$ contre $-4,7 \text{ Kcal.mol}^{-1}$ pour le MES. Nous pourrions ainsi espérer pouvoir stabiliser la nouvelle forme dimérique sur la protéine sauvage avec des dérivés du Raltégravir présentant une très forte affinité pour l'interface (énergie libre de liaison d'au moins $-10 \text{ Kcal.mol}^{-1}$).

Les chimistes avec qui nous collaborons, ont alors développé une seconde série de molécules optimisées à partir du Raltégravir, en ajoutant un noyau de type indène et en enlevant les fonctions dicétone afin que ces composés n'entrent pas en compétition avec le Raltégravir au niveau du site actif. De nombreuses molécules ont alors été imaginées, « dockées » et optimisées sur la base de deux familles d'inhibiteurs potentiels jusqu'à obtenir des énergies libres de liaison prédites de $-10,4 \text{ Kcal.mol}^{-1}$ (correspondant à une affinité de l'ordre du nM à une température de 25 °C) (**figure 36, haut, droite**). Au cours de son stage, N. Aronsohn a synthétisé et isolé plusieurs composés appartenant à la seconde famille et nous les a fournis sous une forme solide (**figure 36, bas**).

Afin de déterminer si ces composés présentaient une affinité pour la nouvelle interface dimérique d'IN-CCD et pour potentiellement obtenir des données structurales sur le « nouvel » assemblage en complexe avec ces molécules, il nous a tout d'abord fallu vérifier si ces composés pouvaient être solubilisés dans des conditions compatibles pour les expériences ultérieures.

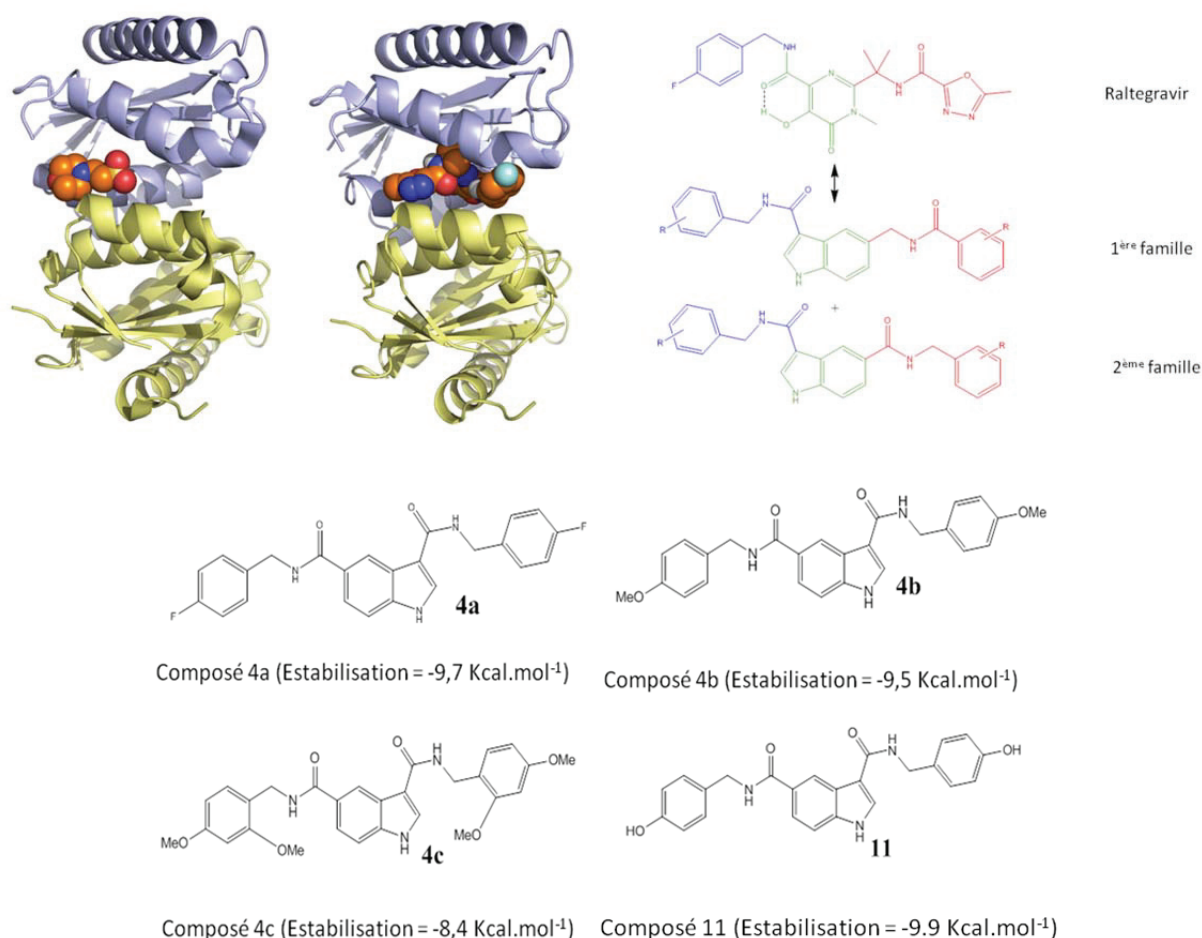


Figure 36 : (Haut, gauche) Représentation de la nouvelle interface d'IN-CCD-WT vue de profil. (Gauche) La molécule de MES localisée à l'interface est représentée en sphères. (Droite) Résultat du *docking* avec le Raltegravir qui est localisé à l'interface et représenté en sphères.

(Haut, droite) Construction des portraits robots des inhibiteurs allostériques potentiels à partir de la structure du Raltegravir.

(Bas) Molécules dérivées du Raltegravir appartenant à la deuxième famille, et dont la synthèse a été réalisée avec leurs énergies de stabilisation prédites.

1.7.2. Solubilisation des composés

Les précédents composés nous ont été fournis sous forme solide. Il nous a donc fallu dans un premier temps évaluer leurs solubilités. Par la simple observation des structures des molécules, il nous a paru que les composés 4a, 4b et 4c seraient probablement peu solubles du fait du manque de fonctions ionisables, ainsi que le composé 11 mais dans une moindre mesure, ce dernier présentant des fonctions hydroxyles phénoliques potentiellement ionisables. La solubilité de ces composés a tout d'abord été évaluée dans l'eau par l'appréciation de la turbidité du milieu. Sans surprise, les composés se sont révélés être extrêmement insolubles dans l'eau même à faible concentration (micromolaire).

Nous avons donc tenté de solubiliser ces composés par l'ajout de dyméthylsulfoxyde (DMSO), qui est un solvant polaire organique, à des concentrations croissantes. Il est apparu que le DMSO était nécessaire pour permettre la solubilisation des composés 4a, 4b et 11. Le composé 4c

présentant quatre fonctions éther de méthyle, est toutefois resté majoritairement insoluble même en présence de DMSO. Ce dernier composé n'a donc pas été retenu pour les expériences ultérieures.

Après plusieurs essais, il est apparu que les composés 4a, 4b et 11 pouvaient être solubilisés à des concentrations respectives de 2,5, 2,5 et 10 mM dans une solution de 50 % V/V DMSO permettant la constitution de solutions mères concentrées. Ces composés ont ensuite pu être dilués au dixième et rester sous forme majoritairement soluble permettant l'obtention de concentrations finales respectives de 0,250, 0,250 et 1 mM dans une solution de 5 % V/V DMSO. Ces concentrations sont bien supérieures aux valeurs de constantes de dissociation calculées à partir des énergies de stabilisation prédites et paraissent donc suffisantes pour les expérimentations ultérieures. Par ailleurs, la concentration finale de DMSO nécessaire au maintien de ces composés sous forme soluble (5 % V/V DMSO) est acceptable et compatible avec la plupart des protéines et des expérimentations de biochimie.

1.7.3. Mesure *in vitro* des interactions entre le nouvel assemblage et les composés synthétisés

Une fois la forme dimérique d'IN-CCD-H103C stabilisée par un pont disulfure, nous avons tenté de mettre en évidence l'interaction en solution du composé MES, ainsi que des composés développées par nos collaborateurs chimistes, avec la nouvelle interface formée. Pour cela, différentes techniques ont été utilisées et parfois mises au point.

1.7.3.A. Mesure par calorimétrie de titration isotherme ou ITC

Un premier essai de mesure d'interaction a été réalisé par calorimétrie de titration isotherme (ITC) entre la forme dimérique stabilisée d'IN-CCD-H103C (30 μ M ou 0,45 mg/ml dans 300 μ l) et le composé MES. Cet essai a été réalisé en collaboration avec Roland Montserret sur un appareil de la plateforme de biophysique de l'IBCP.

Les résultats obtenus n'ont pas indiqué de phénomène d'interaction entre le MES et la forme dimérique stabilisée jusqu'à une concentration testée de MES de 300 μ M. Nous avons par la suite voulu initier de nouvelles expériences d'ITC pour tester des concentrations plus élevées de composé. Cependant les besoins en quantité de protéines se sont avérés être trop importants, ne nous permettant pas de réaliser de nouvelles mesures. Par ailleurs, la technique nous est apparue comme étant incompatible avec le criblage des molécules du fait des concentrations élevées de protéine nécessaires pour chaque essai. Pour ces raisons, nous avons donc recherché des techniques alternatives permettant d'obtenir une information qualitative et quantitative sur les interactions petites molécules/dimère IN-CCD-H103C stabilisé.

1.7.3.B. Mesure par atténuation de fluorescence intrinsèque des protéines

L'observation de la structure cristalline de la nouvelle forme dimérique d'IN-CCD-WT, a permis de mettre en évidence que la liaison du MES au niveau de l'interface se faisait à proximité d'un tryptophane (W138) (**figure 37**). Nous avons donc envisagé une expérience de spectroscopie de fluorescence afin de suivre la fluorescence de la protéine en fonction de la concentration en composé MES ajouté. La proximité du MES avec le résidu W138 était susceptible de modifier (augmenter ou diminuer) la fluorescence de ce tryptophane de la protéine permettant d'envisager une titration du dimère IN-CCD-H103C stabilisé par un pont disulfure avec le MES pour déterminer les paramètres de liaison, comme la constante de dissociation de l'interaction.

Une concentration protéique de 0,01 mg/ml dans un volume de 200 µl de tampon (100 mM Bis-Tris HCl pH 6,0, 500 mM NaCl) était suffisante pour obtenir un bon signal de fluorescence et cette technique s'est donc avérée peu coûteuse en matériel biologique. Le tampon utilisé (Bis-Tris HCl) était concentré (100 mM) afin d'éviter les fluctuations de pH pouvant se produire par l'ajout du composé MES (qui est également un tampon), ces fluctuations pouvant interférer sur la fluorescence des tryptophanes. Le pH choisi pour les expérimentations a été fixé à 6,0, afin de se placer dans une condition de pH identique à celle retrouvée dans la solution de cristallisation d'IN-CCD-WT. Des essais ont donc été réalisés sur le dimère IN-CCD-H103C stabilisé par un pont disulfure et titré par une concentration croissante de MES (jusqu'à 100 mM). Comme contrôle, une solution de NATA (5 µM) a également été titrée avec le MES.

Les mesures réalisées jusqu'à présent n'ont pas permis de mettre en évidence de modification significative de la fluorescence des tryptophanes induite par l'ajout de MES, même pour une concentration élevée (100 mM). Ce résultat peut indiquer l'absence d'interaction du MES sur la forme dimérique IN-CCD-H103C stabilisée et pour les concentrations de MES testées, ou l'absence de modification significative de la fluorescence après saturation de la protéine avec le MES. De plus, des problèmes se sont par la suite posés avec les nouvelles molécules synthétisées qui possèdent des noyaux indènes et induisent un important effet de filtre interne sur la fluorescence des tryptophanes. L'étude par spectroscopie de fluorescence de la liaison des nouveaux composés avec la protéine dimérique IN-CCD-H103C n'a donc pas été poursuivie. Pour cette raison, nous avons tenté de mettre au point une autre technique, peu coûteuse en matériel biologique (protéine), et adaptée à la réalisation de cribles d'interactions.

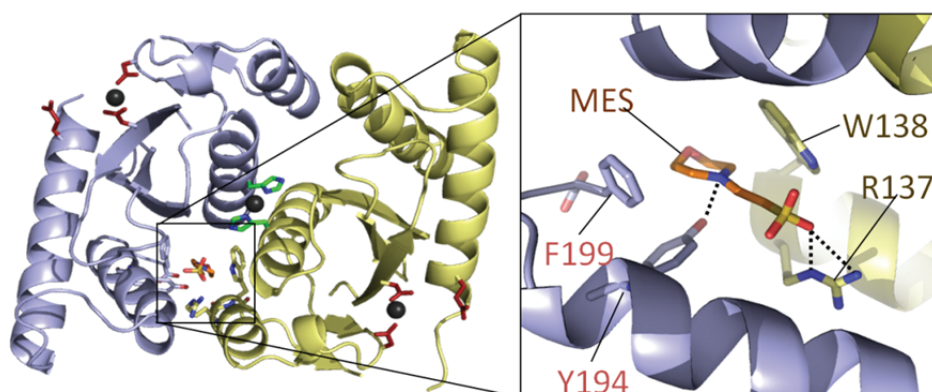


Figure 37 : Vue rapprochée de la molécule de MES enfouie dans l'interface dimérique d'IN-CCD-WT

1.7.3.C. Mesure par fluorimétrie différentielle à balayage thermique

Une autre technique, basée sur l'augmentation de la stabilité d'une protéine en présence d'un ligand, par rapport à la protéine libre, a été tentée. Cette technique appelée fluorimétrie différentielle à balayage thermique ou DSF (pour *Differential Scanning Fluorimetry*), utilise une sonde fluorescente (sonde reportrice) dont la fluorescence est augmentée (effet hyperchrome) lors de la dénaturation thermique d'une protéine, du fait de son interaction avec les régions hydrophobes exposées de la protéine dénaturée. Le suivi de la fluorescence de la sonde en fonction de la température et en présence d'une protéine, permet de déterminer plusieurs paramètres thermodynamiques de stabilité de celle-ci dont la température de demi-dénaturation ou T_m .

Pour cela, nous avons tout d'abord déterminé si la protéine (IN-CCD-H103C stabilisé par un pont disulfure) était adaptée pour ce type d'expérimentations ainsi que les besoins nécessaires en concentrations de protéine et de sonde pour avoir un signal de fluorescence interprétable. Des premiers essais ont donc été réalisés avec des volumes finaux de 20 μ l pour différentes concentrations de protéine (0,25, 0,1 et 0,05 mg/ml final) et différentes concentrations de sonde (Sypro Orange, 45, 5 et 1 X final, 5 % V/V DMSO final) dans une solution fortement tamponnée (100 mM Bis-Tris HCl pH 6,0, 500 mM NaCl). L'utilisation d'un pH acide (bien en dessous du pKa d'ionisation de la fonction thiol des cystéines) présente également l'avantage de limiter la possibilité d'un réarrangement des ponts disulfures (*disulfide bond shuffling*) au cours de la dénaturation de la protéine pouvant aboutir à une rupture des ponts disulfures. Une température de 35 à 95 °C a été balayée et la fluorescence de la sonde suivie, en fonction de la température.

Les premiers résultats obtenus indiquent des profils de dénaturation différents selon les concentrations de protéine et de sonde utilisées (**figure 38**). En effet, on peut observer pour une concentration finale en sonde de 1 X que les courbes de dénaturation ne présentent pas de transitions marquées et ce, quelque soit la concentration protéique testée (**figure 38, haut, gauche**). De la même façon, l'augmentation de la concentration en sonde jusqu'à 5 X final permet d'observer des courbes de dénaturation avec des transitions plus marquées, mais toujours trop faibles pour permettre des mesures précises, en particulier de la température de demi-dénaturation T_m . De plus, une fluorescence importante est déjà observée pour une concentration protéique de 0,1 mg/ml avant l'augmentation de la température, témoignant d'un problème d'agrégation avant le début de la mesure (**figure 38, haut, droite**). Ces concentrations (protéine et sonde) ne semblent donc pas adaptées à la poursuite des études par l'utilisation de cette méthode.

Nous avons alors essayé d'augmenter la concentration en sonde jusqu'à 45 X final et en faisant toujours varier la concentration protéique (**figure 38, bas**). L'augmentation de la concentration en sonde a permis d'améliorer significativement le signal de fluorescence avec l'apparition de courbes de dénaturation présentant une transition marquée, en particulier pour les concentrations protéiques de 0,25 et 0,1 mg/ml, la transition obtenue pour la concentration de 0,05 mg/ml étant encore faible. La courbe de dénaturation obtenue est caractéristique d'une dénaturation selon deux états (replié et dénaturé) et suit une sigmoïde de Boltzmann avec une augmentation progressive de la fluorescence au cours de la dénaturation suivi d'une diminution de cette fluorescence liée à déstabilisation des complexes fluorophores/protéine dénaturée sous l'effet des températures très élevées. Le suivi de la dénaturation semble donc optimal pour une concentration en sonde de 45 X et une concentration protéique de 0,25 mg/ml ne nécessitant au final que peu de matériel biologique pour chaque essai.

Un autre intérêt de cette méthode réside en la possibilité de réaliser jusqu'à 100 mesures simultanément et les longueurs d'ondes d'émission et d'excitation du fluorophore choisi pour

l'expérience (Sypro Orange) sont éloignées de celles des molécules étudiées. De plus, les études d'interaction avec un ligand par cette méthode ne nécessitent pas d'avoir une idée préalable de l'ordre de grandeur de la constante de dissociation de l'interaction mais nécessite uniquement une concentration de ligand considérée comme saturante pour la protéine (cette concentration pouvant correspondre par exemple à la limite de solubilité du composé).

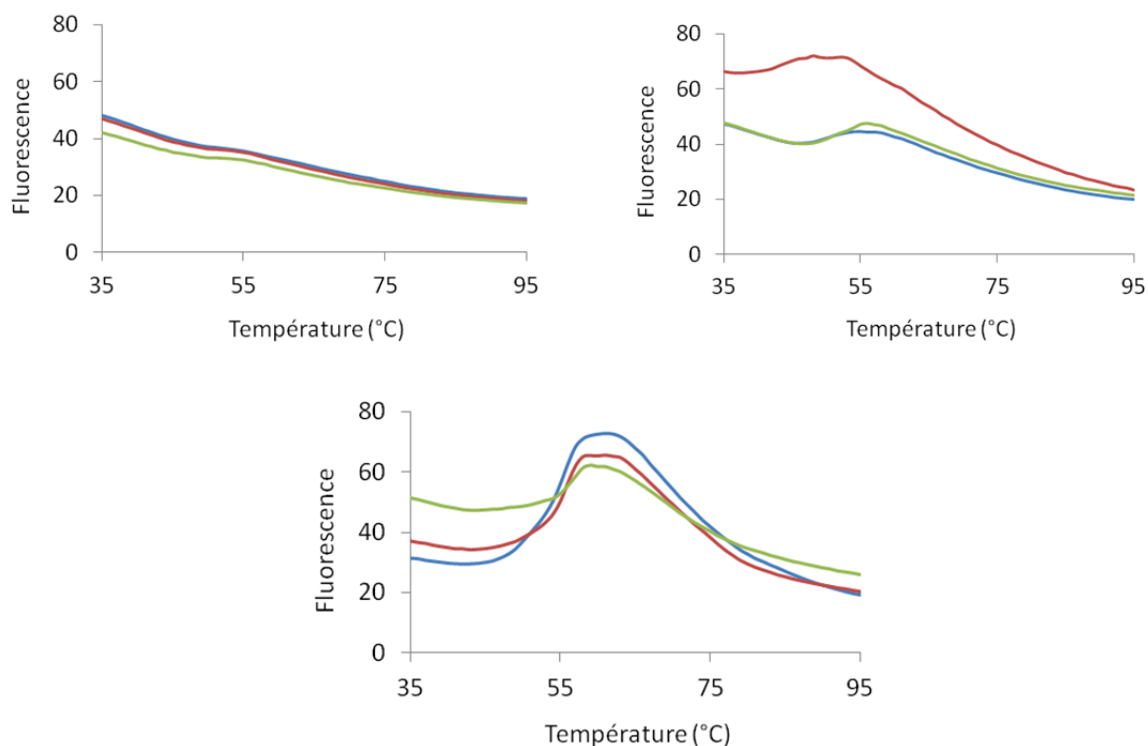


Figure 38 : Courbes de dénaturation obtenues par spectroscopie de fluorescence à balayage thermique (DSF) pour différentes concentrations de protéine et de sonde (Sypro Orange). Axe des ordonnées : fluorescence de la sonde. Axe des abscisses : température en °C. (**Haut, Gauche**) concentration en sonde de 1 X. (**Haut, Droite**) concentration en sonde de 5 X. (**Bas**) concentration en sonde de 45 X. Les courbes vertes, rouges et bleues correspondent respectivement à une concentration protéique de 0,05, 0,1 et 0,25 mg/ml.

Nous avons donc décidé de poursuivre cette étude en comparant les profils de dénaturation obtenus pour IN-CCD-H103C sous forme dimérique, seul, ou incubé avec une forte concentration de MES (100 mM) (**figure 39, gauche**). Si le site de liaison au MES est saturé à cette concentration, nous pourrions nous attendre à une augmentation de la stabilité de la protéine par rapport à la protéine seule. Les courbes de dénaturation indiquent là encore un profil de dénaturation suivant deux états avec une transition bien marquée. Les représentations dérivées premières de ces données donnent un pic pour une température de 60 °C sans le MES et de 60,3 °C avec le MES ce qui correspond aux valeurs des températures de demi-dénaturation (**figure 39, droite**). Cette différence, qui n'est pas supérieur à 0,5 °C, ne peut être considérée comme significative ce qui semble donc indiquer qu'aucun phénomène de fixation saturable ne s'est produit à la concentration de MES testée, corroborant les résultats obtenus par spectroscopie de fluorescence.

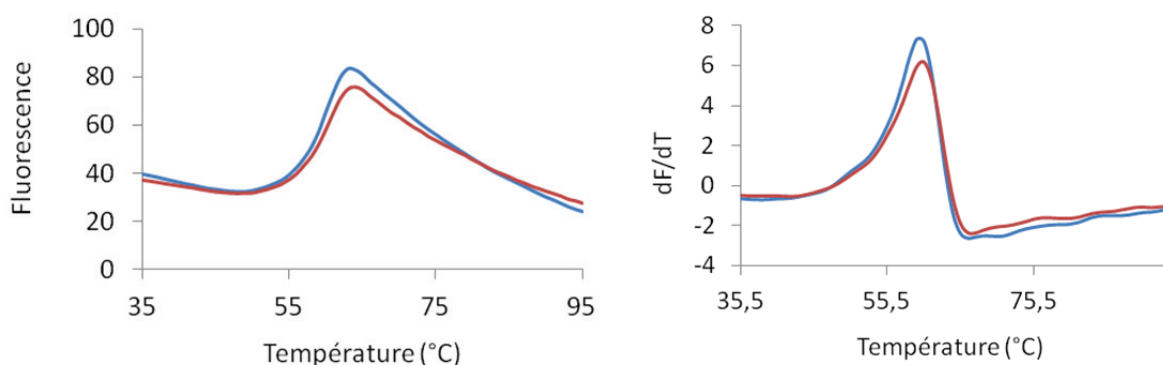


Figure 39 : (**Gauche**) Courbes de dénaturation obtenues par spectroscopie de fluorescence à balayage thermique (DSF) en présence et en absence de MES (100 mM). Axe des ordonnées : fluorescence de la sonde. Axe des abscisses : température en °C. (**Droite**) Courbes dérivées premières. Les courbes bleues et rouges représentent respectivement la protéine seule et la protéine incubée en présence de MES (100 mM).

La molécule de MES étant notre seul contrôle positif pour la fixation de composés à la nouvelle interface de dimérisation, nous avons décidé de tester l'efficacité de cette expérimentation par l'utilisation d'un contrôle positif constitué par un ligand connu de la protéine. En effet, le site actif du domaine catalytique de l'intégrase est capable d'interagir avec différents cations divalents dont le manganèse. Il a été montré qu'une concentration de 10 mM de manganèse était suffisante pour saturer le site actif du domaine catalytique de l'intégrase de RSV (Bujacz *et al.* 1996; Bujacz *et al.* 1997). Nous avons donc décidé de comparer les profils de dénaturation obtenus pour IN-CCD-H103C dimérique, seul ou incubé avec 10 mM de $MnCl_2$ (concentration saturante) ou 10 mM de LiCl en tant que contrôle négatif, les ions monovalents comme le lithium n'étant pas connus pour se fixer dans le site actif des intégrases (**figure 40**).

Les courbes obtenues semblent très similaires entre la protéine seule et la protéine incubée en présence de LiCl. Une différence importante apparaît pour la protéine incubée en présence de $MnCl_2$ avec une augmentation apparente de la stabilité de la protéine (**figure 40, gauche**). Pour obtenir des informations plus précises, les représentations dérivées premières ont été réalisées et les valeurs de températures de demi-dénaturation obtenues sont de 60,3, 60,2 et 63,5 °C respectivement pour la protéine seule, en présence de LiCl et en présence de $MnCl_2$.

L'expérimentation semble donc bien fonctionnelle avec une stabilisation significative de la protéine sous l'effet du manganèse (différence de T_m de 3,2 °C par rapport à la protéine seule) et pouvant être attribuée à la saturation des sites actifs par le magnésium (**figure 40, droite**).

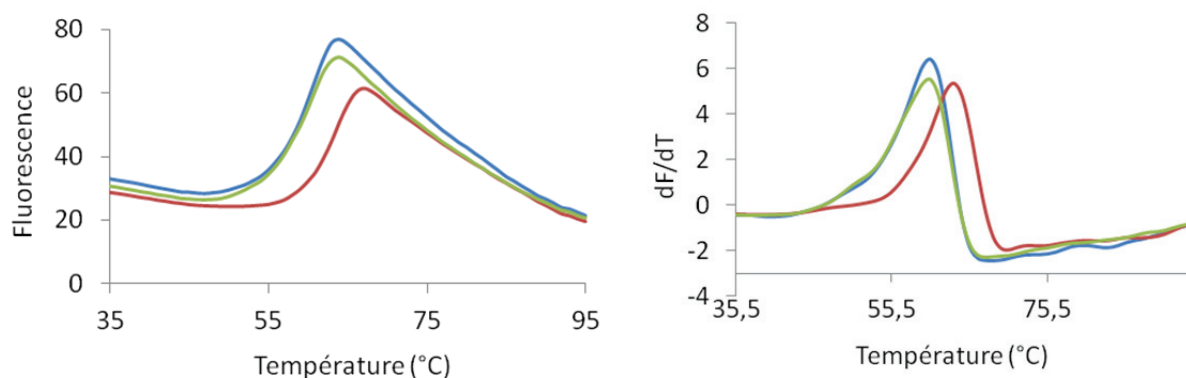


Figure 40 : (**Gauche**) Courbes de dénaturation obtenues par spectroscopie de fluorescence à balayage thermique (DSF) en présence et en absence d'ions. Axe des ordonnées : fluorescence de la sonde. Axe des abscisses : température en °C. (**Droite**) Courbes dérivées premières. Les courbes bleues, vertes et rouges représentent respectivement la protéine seule, en présence de LiCl (10 mM) ou de MnCl₂ (10 mM).

De nouvelles expériences ont donc été réalisées avec le Raltégravir et les composés des chimistes, malgré l'absence de stabilisation de la protéine pour une forte concentration de MES. Les profils de dénaturation obtenus n'ont pas permis d'observer d'augmentation significative de la stabilité d'IN-CCD-H103C sous forme dimérique et pour des concentrations de composés testées d'une centaine de micromolaires, ce qui est bien au dessus des constantes de dissociation prédites par *docking* (données non montrées).

Ainsi, aucune interaction entre les composés et la forme dimérique stabilisée d'IN-CCD-H103C n'a pu être détectée. Nous avons malgré tout tenté de mettre en évidence un effet des composés sur l'état d'oligomérisation d'IN-CCD-WT, qui est majoritairement monomérique en solution. En effet, si ces composés présentent une très forte affinité pour l'interface de dimérisation, ils devraient pouvoir déplacer l'équilibre de la forme monomérique d'IN-CCD-WT dans le sens de la formation de la nouvelle interface.

2. Intégrase IN-CCD-WT

2.1. Effets des composés sur l'état d'oligomérisation d' IN-CCD-WT

Afin de vérifier l'effet des composés sur l'état d'oligomérisation d'IN-CCD-WT, la protéine sauvage a été surproduite et purifiée selon le protocole mis au point par le Dr A. Ballandras au cours de sa thèse. L'étiquette de purification a également été clivée comme pour IN-CCD-H103C.

IN-CCD-WT est monomérique en solution et nous avons voulu vérifier si la protéine pouvait se dimériser en présence des composés synthétisés par les chimistes. Nous avons décidé de vérifier l'effet des composés sur l'état d'oligomérisation d'IN-CCD-WT par la mise en place d'un protocole d'électrophorèse en conditions natives. Un premier protocole a été testé avec un gel de polyacrylamide continu réticulé à 8 % avec un pH de 7,4 maintenu au cours de la migration. Nous avons tout d'abord contrôlé s'il était possible d'observer une différence de migration entre IN-CCD-WT et IN-CCD-H103C sous forme dimérique (après formation du pont disulfure). Pour la migration, les électrodes ont été inversées (de l'anode vers la cathode), IN-CCD possédant un Pi théorique très basique (10,5) et donc bien supérieur au pH utilisé pour l'expérience.

Les échantillons d'IN-CCD-WT et d'IN-CCD-H103C (sous forme dimérique après traitement au peroxyde d'hydrogène) ont ainsi été analysés par électrophorèse sur gel continu de polyacrylamide 8 % en conditions natives (**figure 41**). Il apparaît sur le gel que les deux protéines ont correctement migré dans le gel suite à l'inversion des électrodes, témoignant qu'IN-CCD est bien chargé positivement à un pH proche de la neutralité. Il apparaît également que le profil de migration obtenu pour IN-CCD-WT (**figure 41, pistes 1 et 2**) est différent de celui obtenu pour IN-CCD-H103C sous forme dimérique (**figure 41, pistes 3 et 4**). Ainsi, la bande correspondant à IN-CCD-WT migre plus loin que IN-CCD-H103C. Un contrôle a également été réalisé avec la forme dimérique d'IN-CCD-H103C en présence de 10 mM de DTT et une bande migrant au même niveau que pour IN-CCD-WT a été obtenue (données non montrées). Cependant, une bande migrant au même niveau que IN-CCD-H103C sous forme dimérique était toujours présente et pourrait résulter de la reformation du pont disulfure pendant la migration.

La forme dimérique ayant un profil de migration différent de la forme monomérique, nous avons alors analysé les profils de migrations d'IN-CCD-WT seul, en présence du MES, du Raltégravir ou des composés synthétisés par les chimistes. Après deux heures d'incubation avec ces composés (à température ambiante ou à 4 °C), les échantillons ont été analysés par électrophorèse sur gel continu de polyacrylamide 8 % en conditions natives. Cependant, aucune différence de migration n'est apparue en présence des composés (données non montrées). Il est possible que les composés n'aient pas permis le déplacement de l'équilibre de l'état oligomérique d'IN-CCD-WT vers la nouvelle forme dimérique. Nous n'avons cependant pas pu exclure la possibilité d'une dissociation de la forme dimérique au cours de la migration.

Une approche basée sur l'utilisation d'un agent pontant, comme le BS3 ou le glutaraldéhyde, suivie d'une analyse par électrophorèse sur gel discontinu de polyacrylamide en conditions dénaturantes, a également été tentée afin d'évaluer l'effet des composés sur l'état oligomérique d'IN-CCD-WT. Il est cependant apparu que la réaction de pontage chimique n'était pas optimale sur la forme dimérique d'IN-CCD-H103C avec l'apparition d'agrégats ne pouvant rentrer dans le gel. Du fait de l'absence d'un contrôle positif pour la forme dimérique, ces expériences n'ont pas été poursuivies.

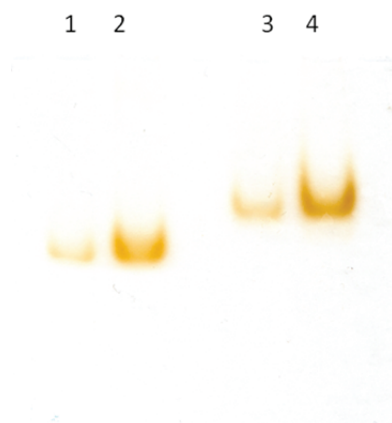


Figure 41 : Analyse par électrophorèse sur gel continu de polyacrylamide (8 %) en conditions natives avec révélation au nitrate d'argent (après une heure de migration à 100 V) d'IN-CCD-WT purifié (piste 1 : dépôt 500 ng, piste 2 : dépôt 1 µg) et IN-CCD-H103C sous forme dimérique (piste 3 : dépôt 500 ng, piste 4 : dépôt 1 µg).

2.2. Trempages des cristaux d'IN-CCD-WT et co-cristallisations avec les composés

Nous avons souhaité réaliser des essais de trempages ainsi que de co-cristallisations des molécules synthétisées avec le CCD sauvage IN-CCD-WT, malgré l'absence de résultats *in vitro* montrant l'interaction du MES, du Raltégavir, ou des autres molécules synthétisées avec le mutant IN-CCD-H103C.

Des cristaux d'IN-CCD-WT ont tout d'abord été obtenus dans les conditions déjà décrites par le Dr A. Ballandras, c'est-à-dire en présence de MES et de zinc. Des essais de trempages ont ensuite été réalisés dans différentes solutions contenant ou non les composés à tester et à différentes concentrations selon leurs limites de solubilités précédemment établies (**tableau 3**). Différents essais ont été réalisés selon les composés avec différents temps de trempages et en présence ou non de MES dans la solution de trempage (le MES étant alors remplacé par un tampon 100 mM Bis-Tris pH 6,0), ce dernier pouvant potentiellement entrer en compétition avec les composés du fait de sa forte concentration dans la solution (100 mM).

Les différents essais réalisés ont tout d'abord mis en évidence que la capacité de diffraction des cristaux était considérablement réduite par l'augmentation des temps de trempage. Ainsi pour le composé 4b, la résolution est passée de 2,5 à 3,25 Å par une augmentation du temps de trempage de seulement trois minutes. Ce phénomène a également été observé pour les trempages réalisés en absence de MES avec une diminution de la résolution encore plus marquée (composés 4b et 4a en absence de MES). Ceci pourrait s'expliquer par la présence de DMSO qui peut induire une dégradation des cristaux pour d'importants temps de trempage mais également pour certains essais par l'absence du MES qui est essentiel à l'empilement cristallin. Il apparaît cependant que des résolutions correctes peuvent être obtenues pour des temps de trempages modérés (entre deux à cinq minutes) et en présence de MES dans la solution.

Les jeux de données enregistrés ont été traités pour remonter à l'information de structure et aux cartes de densité électronique. L'examen des structures pour l'ensemble des trempages réalisés n'ont cependant pas révélé de fixation de ces composés avec une densité électronique correspondant au MES toujours bien présente. Une plus faible occupation du MES a toutefois été observée pour les trempages réalisés en absence de MES.

Des essais de co-cristallisations ont également été réalisés avec ces différents composés mais n'ont pas permis l'obtention de cristaux à ce stade de l'étude.

Composé testé	Concentration	Temps de trempage	Résolution du jeu de données
Raltégravir	0,5 mM	5 min	2,1 Å
		10 min	2,4 Å
		15 min	2,15 Å
		20 min	3,5 Å
Composé 4b	0,25 mM, 5 % V/V DMSO	2 min	2,5 Å
		5 min	3,2 Å
Composé 4b en absence de MES	0,25 mM, 5 % V/V DMSO	2 min	2,9 Å
		5 min	3,55 Å
		10 min	3,9 Å
Composé 4a	0,25 mM, 5 % V/V DMSO	5 min	3,6 Å
Composé 4a en absence de MES	0,25 mM, 5 % V/V DMSO	5 min	3,5 Å
		10 min	4,1 Å
Composé 11	1 mM, 5 % V/V DMSO	5 min	3,2 Å

Tableau 3 : Tableau récapitulatif des essais de trempages réalisés avec l'identité du composé testé, la concentration et les temps trempages ainsi que les meilleures résolutions des jeux de données obtenus.

2.3. Trempages des cristaux d'IN-CCD-WT et co-cristallisations avec des nucléotides

Afin de comprendre le rôle potentiel de la nouvelle interface dans le cycle viral du RAV-1, des études ont également été réalisées par des expériences de trempages et de co-cristallisations avec des nucléotides et désoxyribonucléotides susceptibles de se lier au sillon basique de la « nouvelle » interface d'IN-CCD-WT. Ces expériences ont été réalisées en premier lieu avec des désoxyribonucléotides dNTP. Des nucléotides non hydrolysables (en prévention d'une hydrolyse par l'intégrase *in crystallo*) à noyau purinique entre le phosphate β et γ (cas du composé AMP-PNP) ou α et β (cas du composé AMP-CPP) ont également été testés (**tableau 4**). L'interaction de ces molécules avec le sillon basique reconstitué sur IN-CCD-H103C sous forme dimérique n'a pas été préalablement vérifiée *in vitro* (par gel natif ou DSF). En effet, des interactions entre des nucléotides et le site actif de l'intégrase du VIH-1 ont déjà été rapportées avec en conséquence des risques de faux-positifs (Lipford *et al.* 1994; Mazumder *et al.* 1996).

Il est apparu pour les expériences de trempages et de co-cristallisations avec les composés AMP-CPP et AMP-PNP que ces derniers précipitaient dans la solution de cristallisation. Cette précipitation pouvait résulter de la présence de zinc dans la solution. Les essais de co-cristallisations ont quand même été réalisés mais les trempages ont été effectués dans une solution identique à la solution de cristallisation mais où le zinc (10 mM) était remplacé par du NaCl (200 mM). Les composés n'ont pas précipité dans cette nouvelle solution qui a donc été utilisée. Les essais de trempages et de co-cristallisations ont permis l'obtention de données avec des résolutions correctes (**tableau 4**).

Les jeux de données enregistrés ont ainsi été traités pour remonter à l'information de structure et aux cartes de densité électronique. L'examen des structures pour l'ensemble des trempages réalisés n'ont cependant pas révélé de fixation de ces composés au niveau du sillon basique. Toutefois, pour l'ensemble de ces trempages, nous avons pu observer la présence de deux phosphates au niveau de l'interface et plus précisément du zinc coordonné par les histidines 103 (**figure 42**) sans autre modification significative au niveau de la structure. La possibilité de deux phosphates avec une occupation $\frac{1}{2}$ ne peut pas non plus être écartée. Une densité supplémentaire est présente (**voir Fo-Fc, figure 42**) mais nous n'avons pu en déterminer l'origine. Nous ne savons pas si la présence de ce phosphate ou de ces phosphates a une signification biologique.

Composé testé	Concentration	méthode	Résolution du jeu de données
AMP-CPP	20 mM	co-cristallisation	2,8 Å
AMP-CPP	20 mM	Trempage 2 min	3,05 Å
AMP-PNP	20 mM	co-cristallisation	2,8 Å
AMP-PNP	20 mM	Trempage 2 min	2,7 Å
dNTP	5 mM	Trempage 5 min	3,15 Å

Tableau 4 : Tableau récapitulatif des essais de trempages et de co-cristallisations réalisés avec l'identité du composé testé (nucléotide), la concentration et les temps trempages ainsi que les meilleures résolutions des jeux de données obtenus.

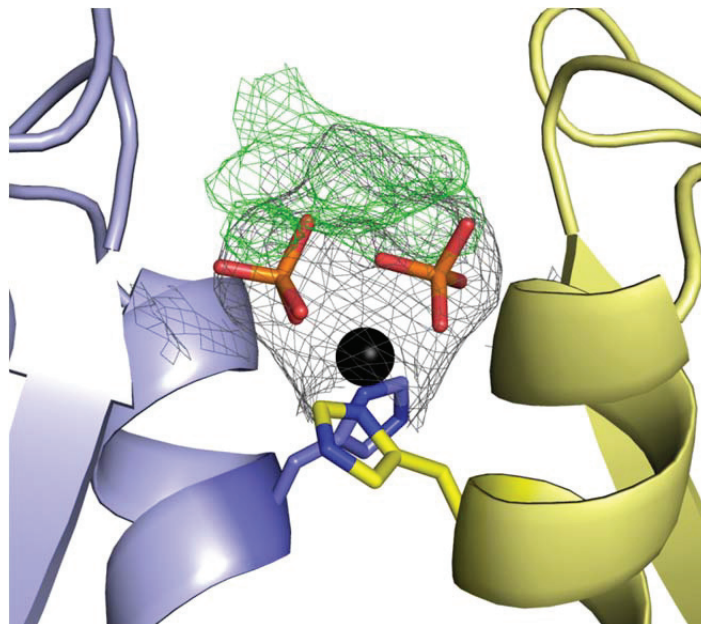


Figure 42 : Cartes de densité électronique (2 Fo-Fc, 1 σ en gris ; Fo-Fc, 3 σ en vert) calculées à partir du jeu de données obtenu avec le trempage AMP-PNP, 20 mM, cinq minutes, 2,7 Å. Deux phosphates sont positionnés (occupation 1,0). Le zinc est représenté par une sphère noire et les histidines 103 de chaque monomère sont également indiquées. Facteur d'agitation thermique moyen de la protéine : 55,3 Å². Facteur d'agitation thermique moyen des phosphates : 62 Å².

3. Conclusion

Un des premiers objectifs visé lors de mon arrivée en thèse était de démontrer que la nouvelle interface de dimérisation, observée pour le domaine catalytique isolé de l'intégrase du RAV-1 ou IN-CCD-WT, n'était pas le résultat d'un artéfact de cristallisation, et était bien capable de se former en solution. Les premières expériences réalisées par le Dr A. Ballandras n'avaient pas permis de démontrer la formation du nouvel assemblage par l'ajout de MES et/ou de zinc. Une stratégie basée sur l'utilisation d'un mutant (H103C) permettant la formation d'un pont disulfure inter-moléculaire a alors été employée.

Nous avons pu montrer par modélisation moléculaire que la formation d'un tel pont disulfure à l'interface du nouvel assemblage dimérique serait possible par la mutation des histidines 103 en cystéines. Le mutant IN-CCD-H103C a été surproduit et purifié dans une souche particulière de *E. coli* (souche origami) adaptée à la formation de ponts disulfures à l'intérieur de la bactérie. L'existence d'une espèce dimérique majoritaire en fin de purification et stabilisée par un pont disulfure entre les cystéines 103 a été confirmée par des analyses de chromatographie d'exclusion de taille, SDS-PAGE (en présence ou absence d'agent réducteur) et de spectrométrie de masse. Cependant, le rendement de purification s'est avéré être très faible (moins de 0,5 mg d'IN-CCD-H103C par litre de culture et avant clivage de l'étiquette de purification).

Le protocole de surproduction a pu être amélioré en utilisant les souches classiques BL21 DE3 pLysS. Cependant, IN-CCD-H103C s'est alors révélé être majoritairement monomérique en fin de purification avec toutefois l'obtention d'un bien meilleur rendement (près de 3 mg de protéine par litre de culture). Un protocole permettant la formation *ex vivo* de la nouvelle interface dimérique sur IN-CCD-H103C et basé sur l'utilisation de peroxyde d'hydrogène, a alors été mis en place pour promouvoir la formation du pont disulfure et a permis l'obtention rapide d'une forme dimérique majoritaire.

Un second objectif consistait à mettre en place des méthodes permettant de mesurer les interactions de petites molécules développées *in silico* et pouvant présenter une affinité élevée pour l'interface de dimérisation. N. Aronssohn (ENS Lyon) qui a réalisé des expériences de *docking* au cours de son stage de Master 1, a abouti à la sélection et à la synthèse de plusieurs candidats. Nous sommes parvenus à solubiliser ces composés en présence de DMSO. L'affinité de ces composés pour l'interface dimérique stabilisée à l'aide du mutant H103C a été évaluée par différentes méthodes (ITC, spectroscopie de fluorescence et fluorimétrie différentielle à balayage thermique) mais aucun phénomène de fixation n'a pu être observé. L'effet de ces composés sur l'état d'oligomérisation d'IN-CCD-WT a également été testé par des expériences de gel natif mais sans succès. Les expériences de trempage et de co-cristallisation n'ont également pas pu mettre en évidence de phénomène de fixation. Toutefois, la fixation de phosphate à l'interface de dimérisation d'IN-CCD-WT a été observée *in crystallo* suite à des essais de trempages avec des nucléotides et désoxyribonucléotides. Nous ne savons cependant pas si leur présence a une signification biologique et nous expliquons difficilement leur présence dans le cas de trempages avec des composés non-hydrolysables tels que l'AMP-CPP et l'AMP-PNP.

En attendant la synthèse de nouveaux composés par nos collaborateurs chimistes de l'ENS de Lyon et de nouveaux essais de trempages avec des acides nucléiques, nous avons décidé de poursuivre nos études sur la protéine entière sauvage dans le but d'obtenir des informations structurales sur cette enzyme (cristallographie et SAXS) et éventuellement de réaliser ces études sur la protéine entière stabilisée selon la nouvelle interface grâce à la création d'un mutant H103C.

1. Intégrase INFLWT-pET-30a

Afin d'initier les études biochimiques et structurales sur l'intégrase entière du RAV-1, sauvage mais aussi mutée (H103C), une stratégie de surproduction en système procaryote (*E. coli*), de solubilisation et de purification de la protéine a été mise en place.

Ces tests ont été réalisés en premier lieu sur la construction INFLWT-pET-30a (contenant le gène codant l'intégrase entière sauvage du RAV-1) qui était disponible dans le laboratoire du Dr C. Ronfort (Moreau *et al.* 2003).

1.1. Surproduction et solubilisation

Les études structurales nécessitent la production de protéine en quantité importante, et son extraction efficace des cellules sous une forme soluble. Pour cela, il est donc nécessaire de disposer d'un protocole efficace de surproduction mais également de solubilisation de la protéine d'intérêt.

Ce protocole a été mis en place pour INFLWT-pET-30a, et l'expression ainsi que la solubilité de la protéine ont été contrôlées. Un premier essai a été réalisé impliquant l'utilisation pour la surproduction de bactéries d'expression BL21 DE3 pLysS transformées par le plasmide pET-30a, qui contient le gène codant l'intégrase entière du RAV-1. Le protocole expérimental comprend un milieu de culture LB avec une induction de la surproduction en phase exponentielle de croissance des bactéries (densité optique à 600 nm comprise entre 0,4 et 0,6) par 1 mM d'IPTG pendant 4 h à 37 °C. Cette induction est suivie d'une lyse des cellules par sonication dans une solution contenant un tampon Tris HCl et du sel (20 mM Tris HCl pH 8,0, 500 mM NaCl et 5 mM β -mercaptoéthanol). Nous avons vérifié la surproduction avec et sans l'inducteur IPTG ainsi que la solubilité de la protéine surproduite (analyse des surnageants ou fractions solubles et des culots ou fractions insolubles) par électrophorèse sur gel discontinu de polyacrylamide 12 % en conditions dénaturantes. La révélation a été réalisée par coloration au bleu de coomassie (**figure 43, gauche**) ou par immuno-empreinte à l'aide d'un anticorps primaire dirigé contre l'étiquette poly-histidines (**figure 43, droite**).

Les résultats obtenus indiquent que l'intégrase INFLWT-pET-30a est bien surproduite suite à l'ajout de l'inducteur IPTG dans le milieu de culture. La protéine surproduite se retrouve majoritairement dans la fraction soluble suite à la lyse des bactéries par sonication (**figure 43, gauche, pistes 1 et 2**). On observe qu'il n'y a que très peu d'intégrase surproduite en absence d'IPTG, ce qui indique l'absence de fuite d'expression significative de notre gène d'intérêt en absence d'inducteur et dans nos conditions de culture (**figure 43, gauche, pistes 3 et 4**). La révélation par immuno-empreinte nous confirme qu'il s'agit bien de la protéine d'intérêt avec l'apparition d'une bande majoritaire migrant au niveau du standard de masse moléculaire de 35 kDa. Cette valeur est proche de la masse moléculaire attendue pour un monomère et calculée à partir de sa séquence protéique (36,7 kDa) (**figure 43, droite, pistes 1 et 2**). Une deuxième bande minoritaire, comprise entre les standards de masses moléculaires de 27 et 35 kDa et peu visible sur le gel coloré au bleu de coomassie, a été révélée par l'anticorps anti-étiquette poly-histidines (**figure 43, droite, piste 2**). Cette bande semble correspondre à un produit de dégradation de l'extrémité carboxy-terminale de la protéine (conservation de l'étiquette poly-histidines) résultant d'une activité de protéolyse résiduelle, malgré la présence d'antiprotéases dans la solution de reprise des culots.

D'autres essais de surproduction ont été réalisés en faisant varier les paramètres de milieu de culture, de température d'induction et de méthode de lyse sans apporter de variations significatives des niveaux d'expression et de solubilité d'INFLWT-pET-30a (données non montrées).

Suite à ces essais, nous avons donc sélectionné les conditions de surproduction et de solubilisation suivantes : une pré-culture de 25 ml de milieu LB supplémenté en kanamycine (50 µg/ml final) et chloramphénicol (50 µg/ml final) est réalisée à partir d'un stock glycérol et incubée à 37 °C sur la nuit sous agitation. Le lendemain, la surproduction de la protéine est réalisée dans un litre de milieu LB supplémenté en kanamycine (50 µg/ml final) à partir de la pré-culture saturée en bactéries. L'induction de la surproduction a lieu par l'ajout d'IPTG (1 mM final) lorsque la culture se trouve approximativement en phase exponentielle de croissance (densité optique à 600 nm comprise entre 0,4 et 0,6) pendant quatre heures à 37 °C. Les bactéries sont ensuite lysées par sonication dans un tampon T1 (20 mM Tris HCl pH 8,0, 500 mM NaCl, 5 mM β-mercaptoéthanol et 20 mM imidazole).

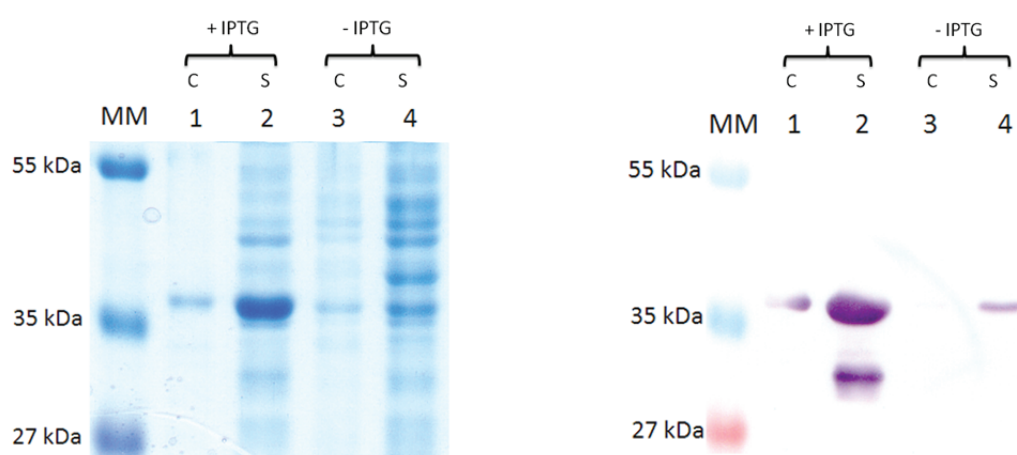


Figure 43 : Analyse par électrophorèse sur gel discontinu de polyacrylamide (12 %) en conditions dénaturantes (SDS-PAGE), de l'expression et de la solubilité de INFLWT-pET-30a. **(Gauche)** Révélation par coloration au bleu de coomassie. **(Droite)** Révélation par immuno-empreinte avec un anticorps primaire anti-étiquette poly-histidines. Piste MM : marqueurs de masses moléculaires connues. Piste 1 : fraction insoluble (témoin induit). Piste 2 : fraction soluble (témoin induit). Piste 3 : fraction insoluble (témoin non induit). Piste 4 : fraction soluble (témoin non induit). C indique la fraction insoluble (culot) et S la fraction soluble (surnageant).

1.2. Purification

Une étape de chromatographie d'affinité pour le nickel a ensuite été réalisée suite à la solubilisation d'INFLWT-pET-30a, afin d'obtenir un échantillon protéique pur et adapté aux études biochimiques et structurales. Après avoir vérifié la présence de notre protéine dans le surnageant, l'intégrase INFLWT-pET-30a qui possède une séquence de six histidines à son extrémité amino-terminale a été purifiée par chromatographie d'affinité pour le nickel à l'aide d'une colonne possédant une phase solide en résine avec un bras espaceur IDA chélatée avec du nickel (colonne Ni-IDA). L'absorbance à 280 nm a été suivie tout le long du processus de purification ainsi que la conductance propre à chaque tampon (**figure 44**).

Après injection de la solution de surnageant sur la colonne et passage de la totalité de la fraction non retenue, un premier lavage dans le tampon T1 (20 mM Tris HCl pH 8,0, 500 mM NaCl, 5 mM β-mercaptoéthanol et 20 mM imidazole) suivi d'un lavage dans un tampon T2 de plus forte force ionique (identique au tampon T1 mais contenant 1 M de NaCl) a permis de décrocher de nombreux

contaminants fixés de façon aspécifique sur la colonne (protéines et acides nucléiques) (**figure 44 fractions A et B**). L'élution de la protéine a ensuite été réalisée grâce à un gradient d'imidazole par mélange du tampon T1 avec un tampon T3 (20 mM Tris HCl pH 8,0, 500 mM NaCl, 5 mM β -mercaptoéthanol et 1 M imidazole). La protéine se décroche pour un gradient 90 % V/V de tampon T1/10 % V/V de tampon T3, ce qui correspond à une concentration de près de 100 mM d'imidazole (**figure 44 fraction C**). L'élution a ensuite été prolongée en maintenant en palier le gradient d'imidazole.

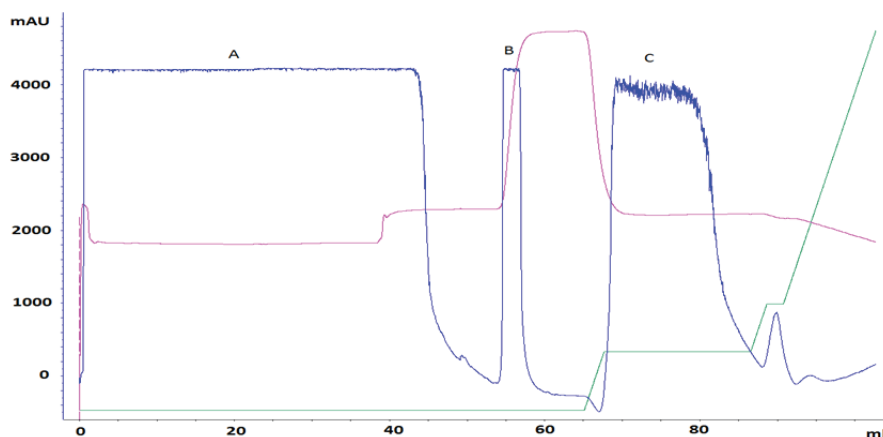


Figure 44 : Chromatogramme de la purification d'INFLWT-pET-30a par chromatographie d'affinité pour le nickel. Courbe bleue : suivi de l'absorbance en sortie de colonne pour une longueur d'onde de 280 nm (axe des ordonnées). Volume d'élution (axe des abscisses). Courbe rose : suivi de la conductance du tampon en sortie de colonne. Courbe verte : suivi du gradient tampon T1/tampon T3 exprimé en pourcentage V/V de tampon T3. Fraction A : fraction non retenue. Fraction B : fraction de lavage. Fraction C : fraction d'élution.

Les fractions non retenues, de lavage et d'élution ont ensuite été analysées par électrophorèse sur gel discontinu de polyacrylamide 12 % en conditions dénaturantes (**figure 45**). La protéine est présente dans les fractions d'élution et est estimée pure à près de 95 % pour les dernières fractions récupérées (**figure 45, gauche, pistes 7 à 13** et **figure 46, droite, pistes 14 à 20**). Les fractions les plus pures ont ensuite été rassemblées puis dialysées contre le tampon T1 sans imidazole (**figure 45, pistes 11 à 20**). Au final, environ 15 mg de protéine par litre de culture ont ainsi été produits, purifiés et stockés à -80 °C dans des cryo-tubes de polypropylène. L'intégrité de la protéine en fin de purification a été évaluée par électrophorèse sur gel discontinu de polyacrylamide 12 % en conditions dénaturantes avec coloration au bleu de coomassie (**figure 46, gauche**) ou révélation au nitrate d'argent (**figure 46, droite**). Les gels nous montrent que notre échantillon présente une haute pureté avec peu de produits de dégradation. L'intégrase purifiée a également été testée pour les activités de clivage et de transfert de brins par le Dr Corinne Ronfort, et est bien fonctionnelle (données non montrées). La protéine étant active en présence de son étiquette de purification, nous avons donc décidé de la conserver pour les expériences ultérieures.

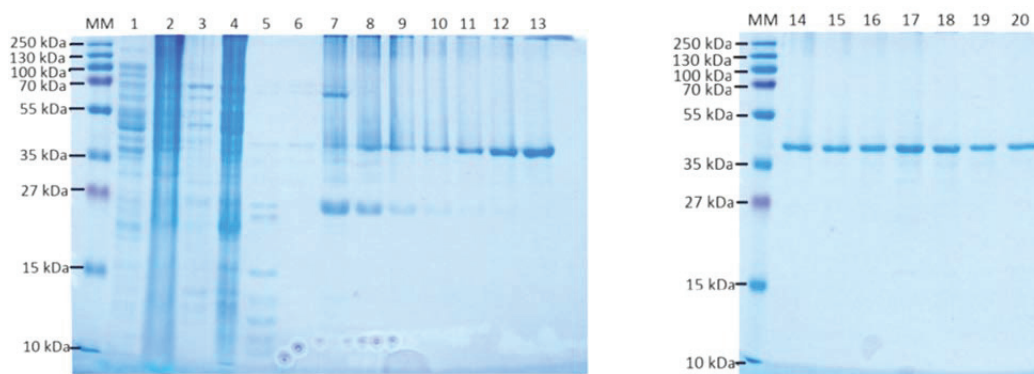


Figure 45 : Analyse par électrophorèse sur gel discontinu de polyacrylamide (12 %) en conditions dénaturantes, de la purification par chromatographie d'affinité pour le nickel d'INFLWT-pET-30a avec révélation par coloration au bleu de coomassie. **(Gauche)** Piste MM : marqueurs de masses moléculaires connues. Piste 1 : fraction insoluble (témoin non induit). Piste 2 : fraction insoluble (témoin induit). Piste 3 : fraction soluble (témoin induit). Piste 4 : fractions non retenues (fraction A sur le chromatogramme). Piste 5 : fractions de lavage (fraction B sur le chromatogramme). Piste 6 à 13 : fractions d'élution (fraction C sur le chromatogramme). **(Droite)** Second gel d'analyse. Piste MM : marqueurs de masses moléculaires connues. Piste 14 à 20 : suite des fractions d'élution.



Figure 46 : Analyse par électrophorèse sur gel discontinu de polyacrylamide (12 %) en conditions dénaturantes, d'INFLWT-pET-30a après décongélation. **(Gauche)** Révélation par coloration au bleu de coomassie. **(Droite)** Révélation au nitrate d'argent. Piste MM : marqueurs de masses moléculaires connues. Piste 1 : fraction purifiée de l'intégrase après décongélation. Piste 2 : fraction purifiée de l'intégrase après décongélation et avec dilution $\frac{1}{2}$.

1.3. Caractérisation de l'échantillon

1.3.1. Analyse par spectrométrie de masse

Une analyse par spectrométrie de masse (MALDI-TOF) a été réalisée sur l'échantillon protéique permettant la détermination de la masse moléculaire d'INFLWT-pET-30a (contrôle de son intégrité moléculaire) ainsi que la détection d'éventuelles modifications chimiques.

Le spectre MALDI-TOF obtenu sur notre échantillon met en évidence la présence d'un ion majoritaire monochargé pour une masse de 36830,5 daltons (**figure 47**). Les ions monochargés constituent les ions majoritaires obtenus suite à une analyse MALDI-TOF et la masse obtenue correspond de très près à celle attendue pour un monomère d'intégrase et calculée à partir de sa séquence protéique (36724,8 Da). Une analyse plus approfondie de la masse obtenue, nous permet de penser que l'ion monochargé pourrait correspondre à un ion $[M+Na]^+$ et semble indiquer une modification d'un thiol libre de l'intégrase par une molécule de β -mercaptoéthanol, présente dans le tampon de stockage de la protéine. La masse moléculaire calculée correspondrait alors à $M = 36724,8 + 23 (Na) + 78,1 (\beta\text{-mercaptoéthanol}) = 36825,9 \text{ Da}$ ce qui est très proche de la masse

moléculaire obtenue pour l'ion monochargé (36830,5 Da). Les ions di-chargés et tri-chargés, qui sont minoritaires, sont également observés respectivement pour les rapports m/z de 18423,9 et 12279,9. Un autre ion minoritaire, pouvant correspondre à un cluster (ion monochargé mais de masse doublée) est également observé dans la région des rapports m/z élevées (73951,9). Il apparaît donc sur cette analyse que l'intégrité moléculaire de l'intégrase est bien conservée et qu'une fonction thiol semble modifiée par une molécule de β -mercaptoéthanol. Une analyse supplémentaire de séquençage par la méthode ESI-MS/MS (spectrométrie de masse en tandem) a par ailleurs permis de confirmer la séquence protéique d'INFLWT-pET-30a (données non montrées).

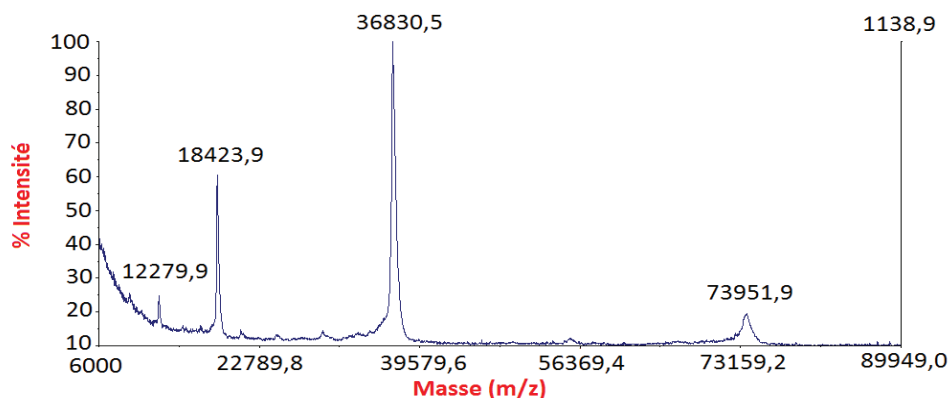


Figure 47 : Analyse par spectrométrie de masse MALDI-TOF d'INFLWT-pET-30a.

1.3.2. Analyse par spectroscopie à corrélation de photons

Les échantillons purifiés d'INFLWT-pET-30a ont été analysés par spectroscopie à corrélation de photons (ou diffusion dynamique de la lumière) à différentes concentrations (gamme de 1 à 3 mg/ml et au-delà de 3 mg/ml) afin de vérifier la monodispersité de l'échantillon et obtenir des informations sur l'état d'oligomérisation de la protéine en fonction de sa concentration (**figure 48**).

Pour toutes les concentrations testées, nous pouvons observer la présence d'un pic sur les profils de distribution du volume caractérisé par un indice de polydispersité proche de 0,2. Ce résultat nous indique que la solution d'INFLWT-pET-30a semble monodisperse. On observe également pour des concentrations protéiques de 1 à 3 mg/ml que la valeur moyenne du diamètre hydrodynamique obtenue sur les profils de distribution du volume et à partir de trois mesures successives est d'environ 8 nm (**figure 48, gauche**). Au delà de 3 mg/ml, la valeur moyenne du diamètre hydrodynamique, obtenue sur les profils de distribution du volume et à partir de trois mesures successives, augmente à 11 nm (**figure 48, droite**). Des valeurs similaires de diamètre hydrodynamique peuvent être observées pour des protéines globulaires standards avec des masses moléculaires respectives de 75 kDa et 150 kDa. Comme le monomère d'INFLWT-pET-30a a une masse moléculaire proche de 37 kDa, ces résultats suggèrent que la protéine pourrait être sous forme dimérique pour la gamme de concentration de 1 à 3 mg/ml et tétramérique au-delà de 3 mg/ml.

Ces résultats suggèrent donc un effet potentiel de la concentration protéique sur l'état d'oligomérisation d'INFLWT-pET-30a. L'augmentation de la concentration protéique pourrait en effet déplacer l'équilibre des formes oligomériques en faveur de la formation de tétramères alors qu'à faible concentration (en dessous de 3 mg/ml), l'état dimérique serait majoritaire.

Toutefois, la variation du diamètre hydrodynamique apparent en fonction de la concentration protéique, peut également être liée à l'augmentation des forces interparticulaires (attractives ou répulsives) qui ne sont plus négligeables à partir d'une certaine concentration protéique. Une augmentation de forces interparticulaires attractives à partir de la concentration d'intégrase de 3 mg/ml, pourrait ainsi être responsable de l'augmentation du diamètre hydrodynamique apparent sans pour autant refléter un changement de l'état d'oligomérisation de la protéine. De plus, la sensibilité de cette méthode ne permet pas toujours de discerner différents états oligomériques présents dans un échantillon.

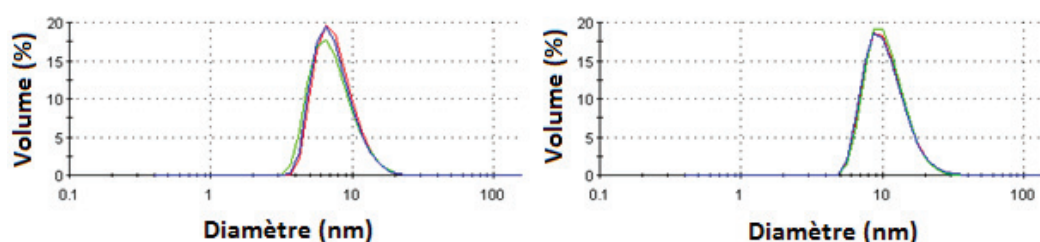


Figure 48 : Profils de distribution du volume (%) en fonction du diamètre hydrodynamique d'INFLWT-pET-30a pour une concentration de 1 à 3 mg/ml (**Gauche**) ou au-delà de 3 mg/ml (**Droite**) obtenus par diffusion dynamique de la lumière (superposition de trois mesures successives).

1.3.3. Analyse par chromatographie d'exclusion de taille

Pour compléter cette étude sur l'état d'oligomérisation d'INFLWT-pET-30a, une expérience de chromatographie d'exclusion de taille a été réalisée. Trois concentrations de protéine ont été testées (1, 3 et 5 mg/ml) et injectées sur une colonne de chromatographie d'exclusion de taille permettant une séparation optimale de macromolécules dont la masse moléculaire est comprise entre 10 et 600 kDa (Superdex 200 HR 10/30) (**figure 49**).

Nous pouvons distinguer sur le profil d'élution obtenu un pic majoritaire qui apparaît au même volume d'élution quelque soit la concentration de protéine injectée. La protéine semble donc en solution présente sous une forme oligomérique majoritaire. La nature de l'état oligomérique de la protéine a été déterminée à partir du volume d'élution correspondant au pic majoritaire et après calibration de la colonne avec des protéines globulaires de masses moléculaires connues. Quelque soit la concentration en protéine injectée, le volume d'élution correspondant au pic majoritaire est voisin de 15,6 ml. Ce volume d'élution est très proche de celui obtenu pour la conalbumine (15,2 ml) dont la masse moléculaire est de 75 kDa. La droite étalon, représentant le logarithme des masses moléculaires des standards en fonction de leurs volumes d'élution, a permis d'estimer une masse moléculaire proche de 65 kDa. Ces valeurs de masses moléculaires sont proches de la masse moléculaire d'un dimère d'intégrase.

Dans ces conditions expérimentales, l'intégrase serait donc majoritairement sous forme dimérique, quelque soit la concentration de protéine injectée. Ces résultats confirment ceux obtenus en DLS, pour lesquels l'intégrase serait sous forme dimérique pour une gamme de concentration de 1 à 3 mg/ml. Toutefois, ils ne permettent pas d'affirmer que l'état oligomérique de l'intégrase serait modifié au-delà de 3 mg/ml.

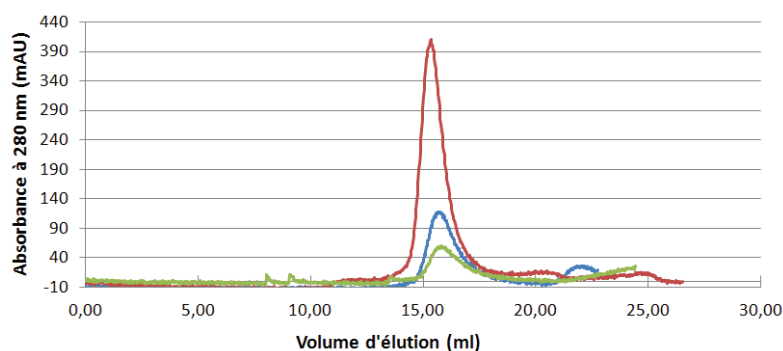


Figure 49 : Profils d'élution d'INFLWT-pET-30a par chromatographie d'exclusion de taille. Courbes vertes, bleues et rouges : suivie de l'absorbance en sortie de colonne pour une longueur d'onde de 280 nm (axe des ordonnées) en fonction du volume d'élution (axe des abscisses) respectivement pour la protéine injectée à 1, 3 et 5 mg/ml.

1.3.4. Analyse par pontage covalent au Bis SulfoSuccinimidyl Suberate (BS3)

Une autre technique a été réalisée afin de caractériser les états d'oligomérisation de l'intégrase en fonction de sa concentration. Une analyse par électrophorèse sur gel discontinu de polyacrylamide 8 % en conditions dénaturantes, a été réalisée après avoir effectué un pontage covalent des protéines grâce à l'agent chimique BS3. Ce pontage doit permettre de figer de manière covalente et irréversible les oligomères présents dans la solution, afin de permettre leur analyse par électrophorèse sur gel de polyacrylamide. Pour cela des solutions d'intégrase à basse concentration (1 mg/ml) et à plus haute concentration (6 mg/ml) ont été soumises au pontage par le BS3 (1 X) puis déposées en quantités équivalentes sur un gel de polyacrylamide (**figure 50**).

Après révélation au nitrate d'argent, on peut observer pour les deux concentrations d'intégrase déposées, la présence d'une bande majoritaire (masse moléculaire proche de 70 kDa) et d'une bande minoritaire (masse moléculaire proche de 35 kDa) (**figure 50, pistes 1 et 2**). Ces résultats suggèrent que dans ces conditions expérimentales, l'intégrase serait majoritairement dimérique quelque soit la concentration testée avec une forme monomérique minoritaire. Ces résultats sont en accord avec une partie de ceux obtenus pour la DLS et la chromatographie d'exclusion pour lesquels l'intégrase serait sous forme dimérique de 1 à 3 mg/ml. Cependant, cette expérience de pontage ne confirme pas pleinement les expériences précédentes. En effet, des formes monomériques sont observées, et pourraient être le résultat d'une réaction incomplète du BS3 du fait de la présence de Tris dans la solution tampon ; ces molécules de Tris peuvent en effet rentrer en compétition avec les fonctions amines de la protéine pour la réaction de pontage. De plus, la présence d'une forme tétramérique au-delà de 3 mg/ml n'est pas confirmée par les expériences de pontage.

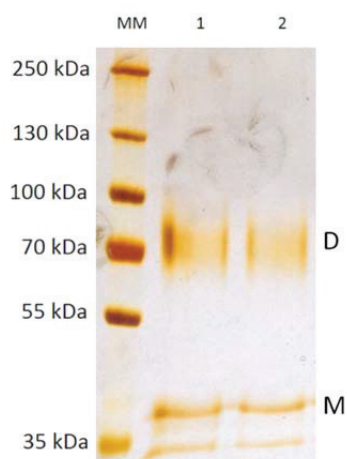


Figure 50 : Analyse de l'état d'oligomérisation d'INFLWT-pET-30a par électrophorèse sur gel discontinu de polyacrylamide (8 %) en conditions dénaturantes et après pontage par le BS3 (1 X) avec révélation au nitrate d'argent. Piste MM : marqueurs de masses moléculaires connues. Piste 1 : dépôt de l'intégrase (1 mg/ml). Piste 2 : dépôt de l'intégrase (6 mg/ml). M et D indiquent respectivement le front de migration de la forme monomérique et dimérique.

Les résultats obtenus par DLS et par chromatographie d'exclusion montrent que notre solution protéique est monodisperse, ce qui est un préalable nécessaire pour pouvoir, d'une part initier des essais de cristallisation, et d'autre part faire une analyse de la protéine en solution par diffusion des rayons X aux petits angles (SAXS). Les résultats de DLS, chromatographie d'exclusion et de pontage BS3 montrent également que l'intégrase est majoritairement dimérique pour une gamme de concentration de 1 à 3 mg/ml. Au delà de cette gamme de concentration, l'état d'oligomérisation reste incertain de part les possibles effets d'attractions interparticulaires.

1.4. Essais de cristallisation

Près de 300 conditions de cristallisation ont été testées pour INFLWT-pET-30a par la technique de diffusion de vapeur en goutte assise. Après plusieurs jours, des cristaux de tailles supérieures à 200 μm sont apparus dans trois conditions de cristallisation différentes (**figure 51**). Le pouvoir de diffraction de ces cristaux a été testé, en utilisant le rayonnement synchrotron de l'ESRF à Grenoble. Sur les trois cristaux testés, il n'y a eu aucune diffraction des rayons X. Ce résultat nous laisse penser que ces cristaux ne sont pas des cristaux de sels, mais des cristaux de protéines. En effet, les cristaux de sels ont généralement un fort pouvoir de diffraction et sont identifiables du fait de leurs petits paramètres de maille. Nous avons fait l'hypothèse que les cristaux testés qui ont une morphologie irrégulière pourraient correspondre à des cristaux faiblement ordonnés de protéine. De nouveaux essais de cristallisation ont été initiés autour des quatre conditions de cristallisation, mais aucun nouveau cristal de protéine n'a pu être observé à ce stade de l'étude.

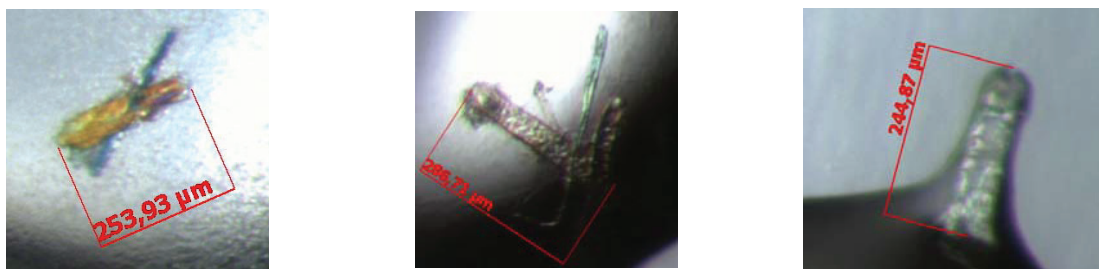


Figure 51 : Cristaux de l'intégrase INFLWT-pET-30a obtenus pour différentes conditions de cristallisation. **(Gauche)** Cristal obtenu avec *Structure Screen 1 et 2 (Molecular Dimension limited)* et dans la condition 0,2 M sulfate d'ammonium, pH 4,6 0,1 M acétate de sodium tri-hydraté, 25 % w/v PEG 4000. **(Milieu)** Cristal obtenu avec *Ammonium Sulfate Suite (Hampton Research)* et dans la condition 0,2 M formate d'ammonium, 2,2 M sulfate d'ammonium. **(Droite)** Cristal obtenu avec *Ammonium Sulfate Suite (Hampton Research)* et dans la condition 0,2 M citrate de di-ammonium, 2,2 M sulfate d'ammonium.

1.5. Diffusion des rayons X aux petits angles (SAXS)

Des mesures de SAXS ont été réalisées pour des solutions d'intégrase à différentes concentrations (0,5, 1, 2 et 3 mg/ml) et pour lesquelles la protéine est majoritairement sous forme dimérique. Des intégrations circulaires ont été réalisées sur les clichés de diffusion obtenus, afin d'obtenir une représentation du logarithme des intensités mesurées en fonction du vecteur de diffusion S .

Les courbes de diffusion obtenues pour les différentes concentrations d'intégrase ont été normalisées après soustraction de la courbe de diffusion obtenue avec le tampon seul (**figure 52, haut**). On peut observer après normalisation, que les courbes correspondant aux solutions concentrées à 0,5 et 1 mg/ml se superposent pour les petits angles de diffusion ($S < 0,040 \text{ \AA}^{-1}$). Pour les concentrations supérieures (2 et 3 mg/ml), on peut observer qu'il n'y a plus de superposition des courbes de diffusion pour les petits angles (**figure 52, bas, gauche**). Cela nous indique que dès 2 mg/ml, des forces attractives entre les particules deviennent non négligeables (décalage des courbes vers le haut) ; pour les concentrations de 0,5 et 1 mg/ml, les forces attractives entre les particules sont encore négligeables du fait des distances qui les séparent. Les informations aux petits angles, qui nécessitent que les forces d'attraction interparticulaires soient négligeables pour avoir de meilleures données sur les particules isolées, ne doivent donc pas être exploitées pour les données obtenues avec les concentrations supérieures à 2 mg/ml, mais peuvent être exploitées pour une concentration de 1 mg/ml.

On observe également que l'augmentation de la concentration entraîne une amélioration du rapport signal/bruit pour les données aux grands angles. Les données à grands angles pour l'intégrase concentrée à 3 mg/ml ont donc été utilisées, ainsi que les données à petits angles pour celle concentrée à 1 mg/ml.

Les rayons de giration ainsi que les masses moléculaires ont été calculés à partir de la droite de Guinier ($S < 0,040 \text{ \AA}^{-1}$) (**figure 52, bas, droite**) pour les solutions à différentes concentrations (tableau 5). On observe que la masse moléculaire et le rayon de giration calculés tendent à augmenter avec la concentration protéique, ce qui témoigne d'un effet d'attraction des particules dépendant de la concentration (tableau 5). La masse moléculaire estimée reste cependant très proche de celle d'un dimère d'intégrase, dont la masse théorique est de 74 kDa. Ceci confirme les résultats précédents obtenus par spectroscopie à corrélation de photons, pour lesquels on observait un dimère d'intégrase pour une concentration de 1 à 3 mg/ml et des forces d'attractions interparticulaires non négligeables au-delà.

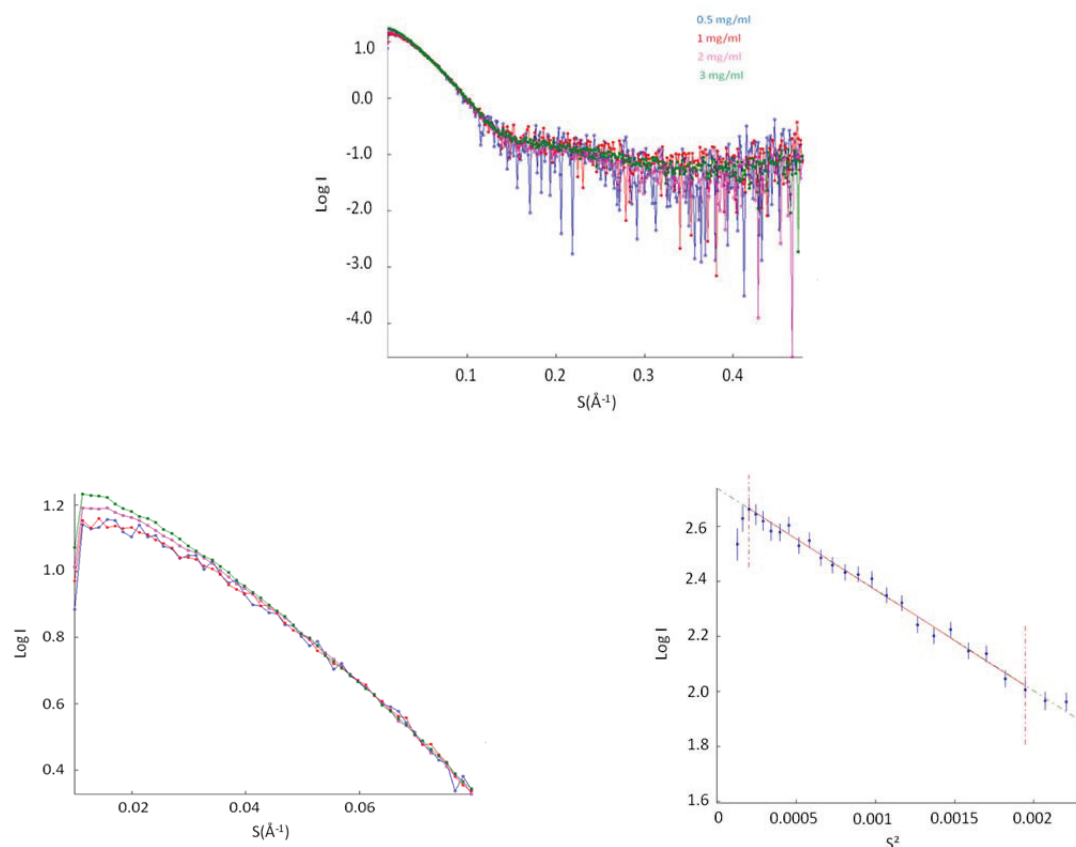


Figure 52 : (Haut, gauche et bas, gauche) Courbes de diffusion des rayons X aux petits angles d'INFLWT-pET-30a en solution représentant le logarithme de l'intensité des rayons X diffusés en fonction du vecteur de diffusion S (\AA^{-1}). Courbe Verte : intégrase à 3 mg/ml. Courbe violette : intégrase à 2 mg/ml. Courbe rouge : intégrase à 1 mg/ml. Courbe bleue : intégrase à 0,5 mg/ml. (Bas, droite) Droite de Guinier représentant le logarithme népérien des intensités diffusées en fonction du vecteur de diffusion S au carré pour l'intégrase à 0,5 mg/ml.

Concentration de l'intégrase (mg/ml)	0,5	1	2	3
Masse moléculaire calculée (kDa)	70,3	69	74	80,9
Rayon de giration calculé (\AA)	33,3	34	35,2	38,3

Tableau 5 : Masses moléculaires et rayons de giration calculés par SAXS pour différentes concentrations d'intégrase.

Le logiciel GNOM, qui calcule la fonction de distribution de distance $P(r)$ par transformée de Fourier indirecte, à partir de la soustraction des courbes de diffusion de la protéine et du tampon permet également de calculer le rayon de giration de la particule étudiée (R_g), ainsi que la distance inter-atomique maximale (D_{max}). La valeur de D_{max} calculée par GNOM est de 110 \AA pour notre intégrase, et est fortement supérieure au diamètre de giration de 66 \AA (figure 53, gauche). Cette différence suggère une forme étendue en solution. La courbe de fonction de distribution de distance $P(r)$ montre également un décalage vers les petites distances interatomiques de son point maximum, qui correspond au R_g , tandis que la représentation de Kratky ($I(S) \cdot S^2 = f(S)$) permet d'observer une courbe en cloche suivie d'un plateau, qui est caractéristique d'une protéine globulaire correctement repliée (figure 54). Un modèle de structure tridimensionnelle de la protéine en solution a ensuite été généré à l'aide du logiciel DAMMIF (figure 53, droite). Pour améliorer le modèle, la courbe de diffusion de la protéine en solution a été reconstruite avec la courbe de diffusion aux petits angles pour la protéine concentrée à 1 mg/ml et la courbe de diffusion aux grands angles pour la protéine concentrée à 3 mg/ml. Le profil de diffusion théorique du modèle en sphères généré par DAMMIF se

superpose correctement au profil de diffusion expérimental. Une symétrie d'ordre 2 peut être observée au sein du dimère d'intégrase. La partie centrale du modèle pourrait correspondre aux domaines catalytiques avec les domaines CTD et NTD de chaque monomère aux deux extrémités.

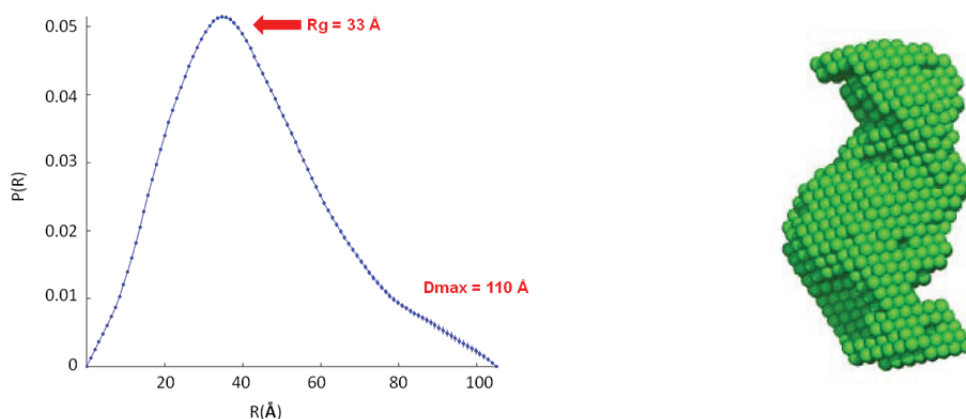


Figure 53 : (Gauche) Courbe de fonction de distribution des distances calculée par GNOM et donnant la distance inter-atomique maximale au sein de la particule (Dmax) ainsi que le rayon de giration (Rg). (Droite) Modélisation de l'enveloppe SAXS du dimère d'INFLWT-pET-30a avec le logiciel DAMMIF.

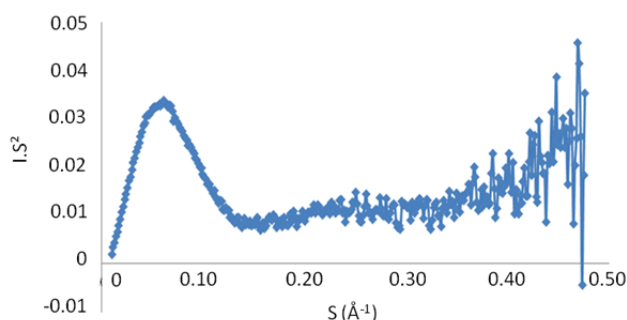


Figure 54 : Représentation de Kratky ($I(S).S^2 = f(S)$) des données de diffusion des rayons X aux petits angles d'INFLWT-pET-30a.

1.6. Conclusion

Un des objectifs de notre étude était d'obtenir des données structurales sur la protéine intégrase (IN) entière du virus associé au virus du sarcome de Rous (RAV-1). Pour cela, nous avons surproduit, purifié et caractérisé l'état d'oligomérisation de l'intégrase INFLWT-pET-30a en fonction de sa concentration, dans l'objectif de lancer des essais de cristallisation et d'obtenir des données structurales par la technique de diffusion des rayons X aux petits angles (SAXS).

La purification d'INFLWT-pET-30a, qui possède une étiquette six histidines à son extrémité amino-terminale, a été réalisée par chromatographie d'affinité pour le nickel. Le protocole sélectionné a permis l'obtention d'un rendement de purification de près de 15 mg de protéine par litre de culture bactérienne. Cette seule étape permet l'obtention d'une protéine estimée pure à près de 95%, après vérification par électrophorèse puis révélation par immuno-empreinte ou coloration

au nitrate d'argent ainsi que par spectrométrie de masse MALDI-TOF. Des études qualitatives sur la protéine en solution ont été réalisées par différentes techniques biophysiques (DLS, SAXS) et biochimiques (tamisage moléculaire, pontage covalent au BS3), afin de déterminer son état d'oligomérisation en fonction de sa concentration. Selon la littérature, l'intégrase peut s'assembler *in vitro* pour former des dimères, des tétramères et parfois des octamères. Les expériences de diffusion dynamique de la lumière nous indiquent que l'intégrase serait majoritairement sous forme dimérique en solution pour une concentration de 1 à 3 mg/ml, en adéquation avec des données disponibles dans la littérature sur cette même enzyme (Jones *et al.* 1992; Coleman *et al.* 1999). Ces résultats ont été confirmés par les expériences de pontage covalent ainsi que par chromatographie d'exclusion. De fortes attractions inter-particulaires apparaissent pour des concentrations au delà de 3 mg/ml.

Des essais de cristallisation ont été initiés et des premiers cristaux ont été obtenus. Ces derniers n'ont cependant pas diffracté les rayons X. Ces cristaux n'ont pas pu être reproduits. Des expériences de SAXS ont été réalisées sur l'intégrase, pour la gamme de concentration où la protéine est dimérique en solution. Les mesures ont été réalisées pour différentes concentrations protéiques (0,5, 1, 2 et 3 mg/ml). La masse moléculaire calculée par cette technique augmente légèrement avec la concentration protéique, mais reste en faveur d'un dimère (masse moléculaire proche de 75 kDa). On ne peut donc pas exclure la présence d'états oligomériques de plus hautes masses moléculaires dans les solutions concentrées de 1 à 3 mg/ml, affectant le calcul de la masse moléculaire par SAXS. Les attractions entre particules sont négligeables pour les concentrations en dessous de 1 mg/ml, mais deviennent importantes au delà de cette concentration. Des données de diffusion des rayons X aux petits angles ont donc pu être obtenues pour la protéine à 1 mg/ml et aux grands angles pour 3 mg/ml. Ces données ont permis la reconstruction d'un modèle tridimensionnel du dimère d'intégrase. Cependant, ce modèle ne permet pas d'identifier clairement l'organisation des domaines de l'intégrase au sein du dimère. La présence de l'étiquette de purification peut également biaiser l'interprétation des résultats de part leur structure désordonnée.

Nous avons donc décidé de cliver l'étiquette de la protéine avant de réaliser de nouvelles expériences de SAXS et pour cela, nous avons réalisé une nouvelle construction plasmidique (construction INFLWT-pPROEX-HTa) avec une étiquette plus courte et clivable par l'endopeptidase du TEV qui est hautement spécifique et facilement surproduite au laboratoire.

2. Intégrase INFLWT-pPROEX-HTa

Afin de réaliser les études biochimiques et structurales pour l'intégrase dans cette nouvelle construction ou intégrase INFLWT-pPROEX-HTa, une stratégie de surproduction en système procaryote (*E. coli*), de solubilisation et de purification de la protéine a été mise en place.

2.1. Surproduction et solubilisation

Suite à l'obtention de la nouvelle construction plasmidique contenant le gène codant l'intégrase du RAV-1 dans le plasmide pPROEX-HTa, nous avons tout d'abord recherché les conditions optimales de culture permettant la surproduction de notre protéine recombinante en grande quantité, ces conditions pouvant être différentes de celles choisies pour l'intégrase INRAV1-pET-30a.

Pour cela, des premiers essais ont été réalisés impliquant l'utilisation pour la surproduction de bactéries Rosetta 2 DE3 pLysS transformées par le plasmide pPROEX-HTa contenant le gène de

l'intégrase du RAV-1, de trois milieux de culture bactériens (milieu LB, Turbo et Hyper) avec une induction de la surproduction (lorsque que la densité optique à 600 nm est proche de 0,6) avec 1 mM d'IPTG ainsi que de deux températures de culture pour la phase d'induction (25 °C sur la nuit ou 37 °C pendant 4 H). Nous avons ensuite vérifié la surproduction avec et sans l'inducteur IPTG par électrophorèse sur gel discontinu de polyacrylamide 12 % en conditions dénaturantes (**figure 55**). La révélation a été réalisée par coloration au bleu de coomassie (**figure 55, gauche**) ou par immuno-empainte à l'aide d'un anticorps primaire dirigé contre l'étiquette poly-histidines (**figure 55, droite**).

Les résultats obtenus indiquent qu'INFLWT-pPROEX-HTa présente des profils d'expression variables selon les paramètres température d'induction et milieu de culture. En présence de l'inducteur IPTG, l'intégrase ne semble pas surproduite pour la température d'induction de 25°C quelque soit le milieu de culture utilisé (**figure 55, gauche et droite, pistes 2, 4 et 6**). Il apparaît cependant que la protéine est bien surproduite pour une température d'induction de 37 °C pour les milieux Hyper et Turbo avec une expression sensiblement plus importante pour le milieu Hyper (**figure 55, gauche et droite, pistes 10 et 12**). En revanche, le milieu LB ne semble pas adapté à la surproduction de la protéine d'intérêt puisque celle-ci n'est pas détectée aussi bien pour les températures d'induction de 25 °C que 37 °C (**figure 55, gauche et droite, pistes 2 et 8**). On observe pour l'ensemble des conditions testées qu'il n'y a que très peu d'intégrase surproduite en absence d'IPTG, ce qui indique l'absence de fuite d'expression de notre gène d'intérêt en absence d'inducteur et dans nos conditions de culture (**figure 55, gauche et droite, pistes 1, 3, 5, 7, 9 et 11**). La révélation par immuno-empainte nous confirme qu'il s'agit bien de la protéine d'intérêt avec l'apparition d'une bande majoritaire migrant au niveau du standard de masse moléculaire de 34 kDa, ce qui correspond de très près à la masse moléculaire théorique d'un monomère d'intégrase calculée à partir de sa séquence protéique (35 kDa) (**figure 55, droite, pistes 10 et 12**).

Suite à ces essais, nous avons donc sélectionné les conditions de surproduction suivantes, avant d'initier les essais de solubilisation : une pré-culture de 20 ml de milieu LB supplémenté en ampicilline (50-100 µg/ml final) et chloramphénicol (50 µg/ml final) est réalisée à partir d'un stock glycérol et incubée à 37 °C sur la nuit sous agitation. Le lendemain, la surproduction de la protéine est réalisée dans un litre de milieu Hyper ou Turbo supplémenté en ampicilline (50-100 µg/ml final) à partir de la pré-culture saturée en bactéries. L'induction de la surproduction a lieu par l'ajout d'IPTG (1 mM final), lorsque la culture présente une densité optique à 600 nm comprise entre 0,4 et 0,6, pendant quatre heures à 37°C.

Un premier essai de solubilisation a été réalisé suite à la surproduction dans les milieux Hyper et Turbo impliquant une lyse des bactéries dans des conditions identiques à celle d'INFLWT-pET-30a (sonication dans une solution contenant un tampon Tris et du sel : 20 mM Tris HCl pH 8,0, 500 mM NaCl et 5 mM β-mercaptoéthanol). Nous avons vérifié la solubilité de la protéine (analyse des surnageants ou fractions solubles et des culots ou fractions insolubles) par électrophorèse sur gel discontinu de polyacrylamide 12 % en conditions dénaturantes (**figure 56**). Les résultats obtenus indiquent que dans des conditions de lyse similaires à celles utilisées pour INFLWT-pET-30a, INFLWT-pPROEX-HTa est retrouvée en très grande majorité dans la fraction insoluble (**figure 56, pistes 1 et 3**). La comparaison des profils de solubilité obtenus pour les deux milieux de culture indique que la protéine reste majoritairement insoluble quelque soit le milieu de culture utilisé.

Nous avons donc cherché à comprendre l'origine de cette insolubilité afin de déterminer des conditions optimales permettant l'extraction de notre protéine d'intérêt. L'insolubilité de notre protéine peut avoir pour origine aussi bien les conditions de surproduction (milieu de culture, température d'induction, concentration d'IPTG, force des promoteurs, souche bactérienne), que les conditions de lyse (solution tampon utilisée, salinité de la solution, pH, méthodes de lyse). Différents paramètres ont ainsi du être modifiés afin de favoriser la solubilisation d'INFLWT-pPROEX-HTa.

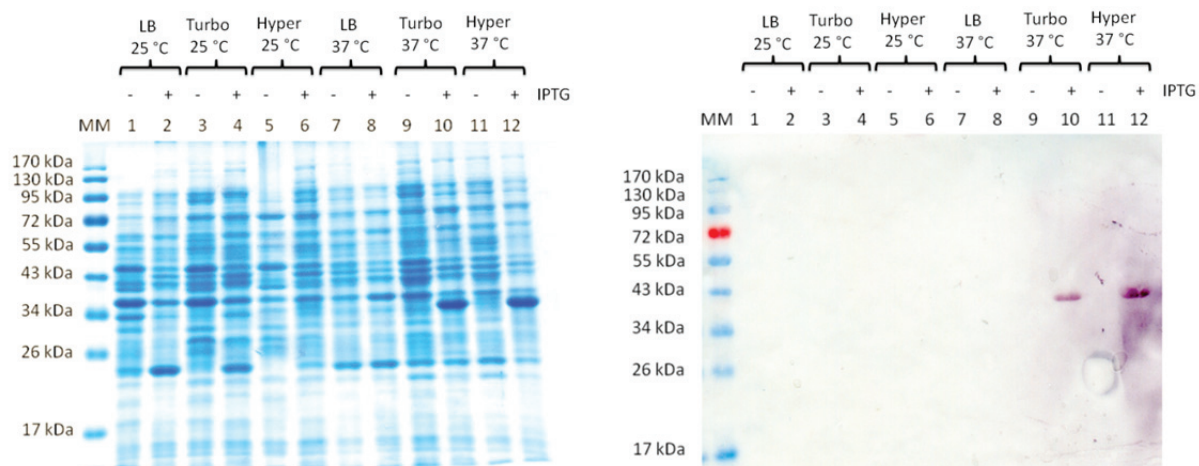


Figure 55 : Analyse par électrophorèse sur gel discontinu de polyacrylamide (12 %) en conditions dénaturantes, de la surproduction d'INFLWT-pPROEX-HTa. **(Gauche)** Révélation par coloration au bleu de coomassie. **(Droite)** Révélation par immuno-empainte avec un anticorps primaire anti-étiquette poly-histidines. Piste MM : marqueurs de masses moléculaires connues. Piste 1 : culture sans IPTG (milieu LB, 25 °C). Piste 2 : culture avec IPTG (milieu LB, 25 °C). Piste 3 : culture sans IPTG (milieu Turbo, 25 °C). Piste 4 : culture avec IPTG (milieu Turbo, 25 °C). Piste 5 : culture sans IPTG (milieu Hyper, 25 °C). Piste 6 : culture avec IPTG (milieu Hyper, 25 °C). Piste 7 : culture sans IPTG (milieu LB, 37 °C). Piste 8 : culture avec IPTG (milieu LB, 37 °C). Piste 9 : culture sans IPTG (milieu Turbo, 37 °C). Piste 10 : culture avec IPTG (milieu Turbo, 37 °C). Piste 11 : culture sans IPTG (milieu Hyper, 37 °C). Piste 12 : culture avec IPTG (milieu Hyper, 37 °C).

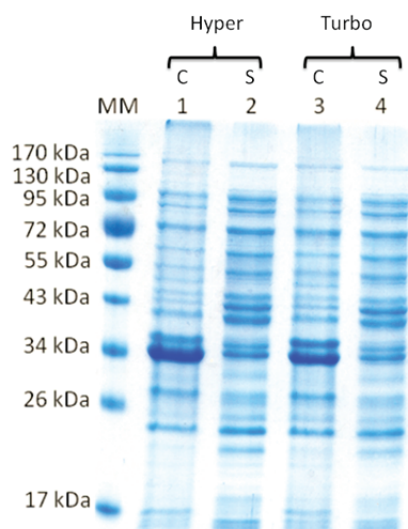


Figure 56 : Analyse par électrophorèse sur gel discontinu de polyacrylamide (12 %) en conditions dénaturantes de la solubilité d'INFLWT-pPROEX-HTa avec révélation par coloration au bleu de coomassie. Piste MM : marqueurs de masses moléculaires connues. Piste 1 : fraction insoluble (milieu Hyper). Piste 2 : fraction soluble (milieu Hyper). Piste 3 : fraction insoluble (milieu Turbo). Piste 4 : fraction soluble (milieu Turbo). C indique la fraction insoluble (culot) et S la fraction soluble (surnageant).

2.2. Essais de solubilisation

Les premiers essais de solubilisation ont mis en évidence que INFLWT-pPROEX-HTa était bien surproduit dans les conditions de culture précédemment établies, mais qu'elle était présente majoritairement dans la fraction insoluble suite à la lyse des bactéries. Afin de tenter d'améliorer la

solubilisation de la protéine recombinante, différents paramètres ont tout d'abord été modifiés pour la culture des bactéries et pour l'induction de la surproduction. Du fait de l'absence de surproduction de notre protéine à la température de 25 °C, nous avons décidé de ne pas modifier le paramètre température d'induction et de rester à 37 °C pour les prochains essais.

Nous avons donc dans un premier temps testé différentes concentrations d'IPTG (0,3 ou 1 mM), ainsi que l'ajout d'un inducteur de la surproduction de chaperonnes chez *E. coli* (Augmedium) ou la mise en place d'un choc thermique à 42 °C durant toute la culture et avant l'induction de la surproduction à 37 °C afin là aussi, d'activer la surproduction de certaines chaperonnes (Chen *et al.* 2002). Il est connu que la réduction de la concentration en IPTG peut avoir des effets bénéfiques pour la solubilité des protéines recombinantes, ainsi que l'augmentation de la surproduction des chaperonnes, qui peut aider au bon repliement et ainsi, à la solubilité des protéines surproduites. L'ensemble de ces paramètres a été testé aussi bien pour le milieu de culture Hyper que le milieu de culture Turbo. Nous avons ensuite vérifié la surproduction avec l'inducteur IPTG (**figure 57, gauche**) et la solubilité (**figure 57, droite**) de la protéine (analyse des surnageants ou fractions solubles et des culots ou fractions insolubles) par électrophorèse sur gel discontinu de polyacrylamide 12 % en conditions dénaturantes.

L'analyse SDS-PAGE nous indique là encore que la surproduction de la protéine est plus importante dans le milieu Hyper que dans le milieu Turbo, quelque soit les paramètres de culture utilisés (**figure 57, gauche, comparaison des pistes 1 à 6 avec 7 à 12**). Nous pouvons observer que la quantité de protéine visible sur le gel est identique pour les deux concentrations d'IPTG testées et pour un même milieu de culture (**figure 57, gauche, comparaison des pistes 1 à 3 avec 4 à 6 et 7 à 9 avec 10 à 12**). La surproduction de notre protéine semble donc déjà maximale pour la concentration de 0,3 mM. Nous pouvons de plus observer que l'ajout d'Augmedium dans le milieu de culture Hyper semble augmenter la surproduction pour les deux concentrations d'IPTG testées (**figure 57, gauche, comparaison de la piste 2 avec les pistes 1 et 3 et de la piste 5 avec les pistes 4 et 6**). Le choc thermique ne semble pas améliorer significativement la quantité de protéine pour le milieu Hyper par rapport au contrôle avec IPTG seul, mais l'améliore sensiblement pour le milieu Turbo, là où l'Augmedium ne semble avoir que peu d'effet.

Nous avons par la suite analysé la solubilité de notre protéine surproduite dans les différentes conditions de culture (**figure 57, droite**). L'analyse SDS-PAGE (représentant uniquement les tests pour le milieu Hyper) présente des profils de solubilité très similaires quelque soit les conditions de culture utilisées. Les mêmes résultats ont été obtenus pour le milieu Turbo. La protéine reste présente majoritairement dans la fraction insoluble et aucune amélioration ne semble à noter par rapport au contrôle IPTG seul. Il apparaît donc que la surproduction de chaperonnes avant l'induction de la surproduction de notre protéine, ainsi que la réduction de la concentration en IPTG utilisée ne permettent pas d'améliorer significativement la solubilité d'INFLWT-pPROHEX-HTa.

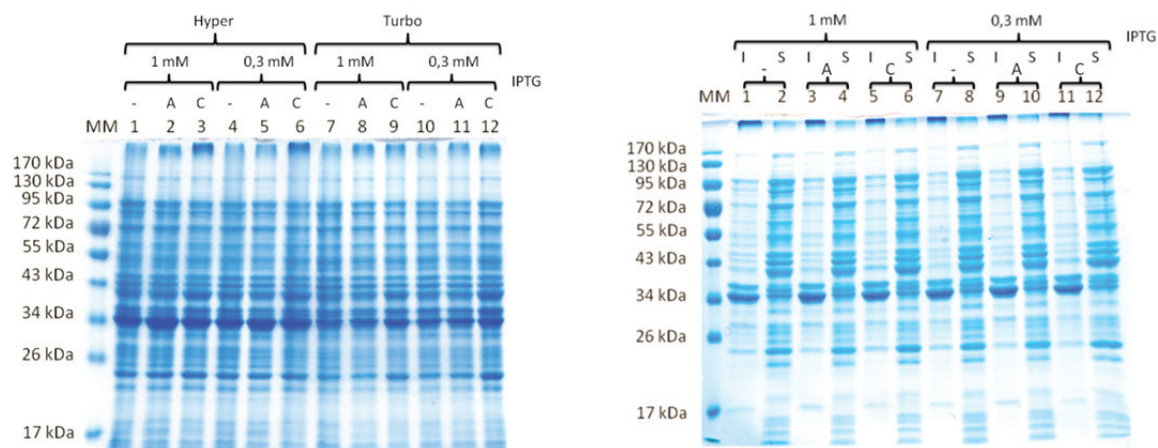


Figure 57 : (Gauche) Analyse par électrophorèse sur gel discontinu de polyacrylamide (12 %) en conditions dénaturantes, de la surproduction d'INFLWT-pPROEX-HTa avec révélation par coloration au bleu de coomassie. Piste MM : marqueurs de masses moléculaires connues. Piste 1 : culture Hyper avec IPTG (1 mM). Piste 2 : culture Hyper avec IPTG (1 mM) et en présence d'Augmedium. Piste 3 : culture Hyper avec IPTG (1 mM) et choc thermique. Piste 4 : culture Hyper avec IPTG (0,3 mM). Piste 5 : culture Hyper avec IPTG (0,3 mM) et en présence d'Augmedium. Piste 6 : culture Hyper avec IPTG (0,3 mM) et choc thermique. Piste 7 : culture Turbo avec IPTG (1 mM). Piste 8 : culture Turbo avec IPTG (1 mM) et en présence d'Augmedium. Piste 9 : culture Turbo avec IPTG (1 mM) et choc thermique. Piste 10 : culture Turbo avec IPTG (0,3 mM). Piste 11 : culture Turbo avec IPTG (0,3 mM) et en présence d'Augmedium. Piste 12 : culture Turbo avec IPTG (0,3 mM) et choc thermique. (—) indique le contrôle IPTG seul, « A » l'essai Augmedium et « C » l'essai choc thermique.

(Droite) Analyse par électrophorèse sur gel discontinu de polyacrylamide (12 %) en conditions dénaturantes, de la solubilité (milieu Hyper) d'INFLWT-pPROEX-HTa avec révélation par coloration au bleu de coomassie. Piste MM : marqueurs de masses moléculaires connues. Pistes 1, 3, 5, 7, 9, 11 : fractions insolubles respectivement pour 1 mM IPTG, 1 mM IPTG et Augmedium, 1 mM IPTG et choc thermique, 0,3 mM IPTG, 0,3 mM IPTG et Augmedium, 0,3 mM IPTG et choc thermique. Pistes 2, 4, 6, 8, 10, 12 : fractions solubles respectivement pour 1 mM IPTG, 1 mM IPTG et Augmedium, 1 mM IPTG et choc thermique, 0,3 mM IPTG, 0,3 mM IPTG et Augmedium, 0,3 mM IPTG et choc thermique. — indique le contrôle IPTG seul, A l'essai Augmedium et C l'essai choc thermique. I et S indiquent respectivement les fractions insolubles et solubles

Suite aux premiers essais pour tenter de solubiliser notre protéine d'intérêt, nous avons décidé de tester le paramètre souche bactérienne pouvant affecter la solubilité d'une protéine surproduite. Pour cela, nous avons testé la surproduction et la solubilité d'INFLWT-pPROEX-HTa dans quatre souches bactériennes différentes, la souche Rosetta 2 DE3 pLysS (contrôle), BL21 DE3 pLysS, C41 DE3 pLysS et C43 DE3 pLysS. Ces différentes bactéries ont été transformées par le plasmide d'intérêt et des essais de surproduction ont été réalisés dans les conditions optimales déterminées pour les milieux Hyper et Turbo mais en réduisant la concentration d'IPTG à 0,3 mM final. Du fait des profils de solubilité similaires obtenus pour les deux milieux de culture, nous avons décidé de poursuivre les essais sur le milieu présentant le niveau de surproduction le plus élevé c'est-à-dire le milieu Hyper. Nous avons ensuite vérifié la surproduction avec et en absence de l'inducteur IPTG et la solubilité de la protéine (analyse des surnageants ou fractions solubles et des culots ou fractions insolubles) par électrophorèse sur gel discontinu de polyacrylamide 12 % en conditions dénaturantes (**figure 58**).

L'analyse SDS-PAGE indique que la protéine est bien surproduite quelque soit la souche utilisée mais également l'absence de fuite d'expression en absence d'inducteur (**figure 58, gauche**). Il apparaît également que les profils de surproduction d'INFLWT-pPROEX-HTa sont très similaires pour les quatre souches bactériennes testées. Nous avons par la suite analysé la solubilité de notre protéine dans les quatre souches utilisées. Là encore, les profils de solubilité sont identiques et ne semblent pas indiquer d'amélioration significative de la solubilité d'INFLWT-pPROEX-HTa (**figure 58,**

droite). L'ensemble des résultats obtenus ne semble donc pas indiquer d'amélioration significative de la solubilité de notre protéine par la modification des paramètres de culture et d'induction. Il nous est donc apparu nécessaire de modifier les conditions d'extraction (solutions tampons utilisées, méthodes de lyse) afin de tenter d'extraire notre protéine de la fraction insoluble.

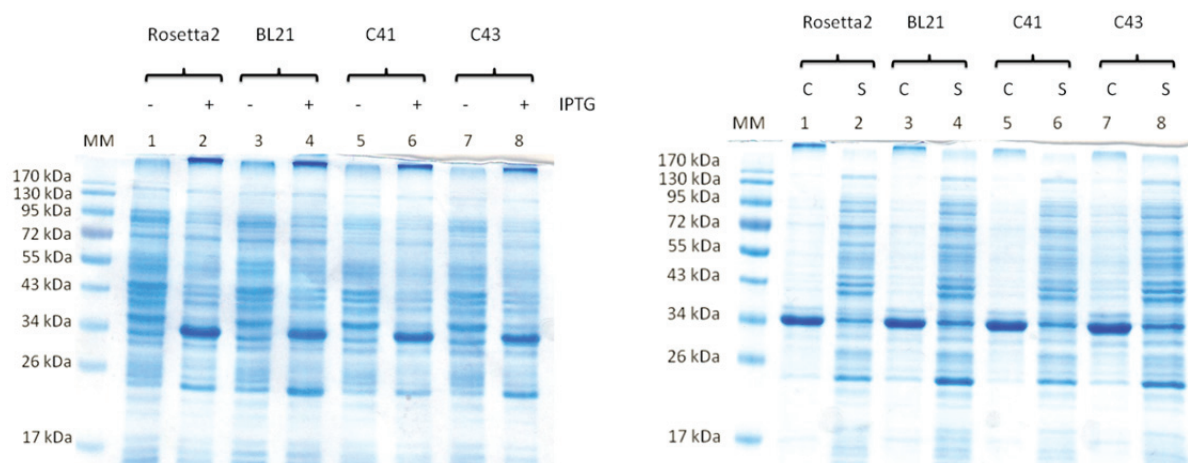


Figure 58. (Gauche) Analyse par électrophorèse sur gel discontinu de polyacrylamide (12 %) en conditions dénaturantes, de la surproduction d'INFLWT-pPROEX-HTa avec révélation par coloration au bleu de coomassie. Piste MM : marqueurs de masses moléculaires connues. Piste 1 : souche Rosetta 2 DE3 pLysS sans IPTG. Piste 2 : souche Rosetta 2 DE3 pLysS avec IPTG. Piste 3 : souche BL21 DE3 pLysS sans IPTG. Piste 4 : souche BL21 DE3 pLysS avec IPTG. Piste 5 : souche C41 DE3 pLysS sans IPTG. Piste 6 : souche C41 DE3 pLysS avec IPTG. Piste 7 : souche C43 DE3 pLysS sans IPTG. Piste 8 : souche C43 DE3 pLysS avec IPTG.

(Droite) Analyse par électrophorèse sur gel discontinu de polyacrylamide (12 %) en conditions dénaturantes, de la solubilité (milieu Hyper) d'INFLWT-pPROEX-HTa avec révélation par coloration au bleu de coomassie. Piste MM : marqueurs de masses moléculaires connues. Pistes 1, 3, 5, 7 : fractions insolubles respectivement pour les souches Rosetta 2 DE3 pLysS, BL21 DE3 pLysS, C41 DE3 pLysS et C43 DE3 pLysS. Pistes 2, 4, 6, 8 : fractions solubles respectivement pour les souches Rosetta 2 DE3 pLysS, BL21 DE3 pLysS, C41 DE3 pLysS et C43 DE3 pLysS. C indique la fraction insoluble (culot) et S la fraction soluble (surnageant).

Différentes méthodes de lyse ont dans un premier temps été testées sans pour autant améliorer l'extraction de notre protéine (données non montrées). Nous avons donc modifié la composition de la solution d'extraction et, après de nombreuses tentatives, une condition de tampon a été déterminée permettant une extraction efficace de notre protéine d'intérêt. La solution d'extraction optimisée possède la composition suivante : 50 mM Bis-Tris HCl pH 7,0, 1 M NaCl, 10 % V/V glycérol, 25 mM imidazole et 1 M urée. L'utilisation d'une forte concentration en sel (1 M) en absence d'urée a permis d'améliorer la solubilité de la protéine mais celle-ci s'est révélée très vite instable avec l'apparition d'une précipitation irréversible. L'urée est apparue comme essentielle pour une extraction efficace d'INFLWT-pPROEX-HTa. L'utilisation d'un agent chaotrope comme l'urée à faible concentration, est non dénaturante pour la protéine et indique que l'intégrase surproduite n'est pas présente dans les corps d'inclusion au cours de la surproduction.

Au cours de l'optimisation des conditions d'extraction, nous avons poursuivi l'optimisation des conditions de surproduction. Les conditions de surproduction et de solubilisation finalement choisies pour INFLWT-pPROEX-HTa sont les suivantes : à partir du stock glycérol, un étalement sur une boîte de pétri LB-agar supplémenté en ampicilline à forte concentration (200 µg/ml), chloramphénicol (50 µg/ml) et glucose (1 % W/V) est réalisé et placé à l'étuve à 37 °C sur la nuit. Le lendemain, les clones obtenus sur une boîte sont resuspendus dans un peu de milieu LB et utilisés pour inoculer 250 ml de milieu de culture LB supplémenté en ampicilline (200 µg/ml), glucose (1 % W/V) et 20 mM Tris HCl pH 7,0 (*Plating method*). L'induction de la surproduction a lieu par l'ajout

d'IPTG (0,3 mM final) et de ZnCl_2 (50 μM final) lorsque la culture présente une densité optique à 600 nm comprise entre 0,8 et 1 pendant quatre heures à 37°C. Pour la solubilisation d'INFLWT-pPROEX-HTa, les bactéries sont resuspendues dans la solution d'extraction ou tampon T1 (50 mM Bis-Tris HCl pH 7,0, 1 M NaCl, 10 % V/V glycérol, 25 mM imidazole et 1 M urée) puis lysées par le système microfluidique.

2.3. Purification

Une étape de chromatographie d'affinité a été réalisée suite à la solubilisation d'INFLWT-pPROEX-HTa, afin d'obtenir un échantillon protéique pur et adapté aux études biochimiques et structurales. A titre de comparaison, une purification a également été réalisée en utilisant les conditions de lyse déterminées pour INFLWT-pET-30a.

Comme pour INFLWT-pET-30a, INFLWT-pPROEX-HTa a été purifiée par chromatographie d'affinité pour le nickel à l'aide d'une colonne possédant une phase solide en résine avec un bras espaceur IDA et chélatée avec du nickel (colonne Ni-IDA). L'absorbance à 280 nm a été suivie tout le long du processus de purification ainsi que la conductance propre à chaque tampon (**figure 59**).

Après injection de la solution de surnageant sur la colonne et passage de la totalité de la fraction non retenue, une solution identique au tampon de lyse, mais avec 0,5 M NaCl (tampon T2), est injectée sur la colonne puis la concentration en sel est à nouveau augmentée à 1 M avec le tampon de lyse. Ce choc ionique ne semble pas avoir décroché de contaminants de la colonne, probablement du fait de la concentration en sel déjà très élevée dans le tampon de lyse. L'élution de la protéine a ensuite été réalisée grâce à un gradient d'EDTA par mélange du tampon T1 avec un tampon T3 (50 mM Bis-Tris HCl pH 7,0, 1 M NaCl, 10 % V/V glycérol, 1 M urée, 0,1 M EDTA). Il est en effet apparu au cours de nos essais que l'élution classique en utilisant l'imidazole aboutissait très rapidement en fin de purification à une protéine agrégée. Il est également apparu qu'il était nécessaire de conserver l'urée jusqu'aux dernières étapes de purification. En utilisant l'élution par l'EDTA, notre protéine semble se décrocher pour un gradient 50 % V/V de tampon T1/50 % V/V de tampon T3 ce qui correspond à une concentration de 50 mM d'EDTA. L'élution a ensuite été prolongée sans maintenir de palier jusqu'à atteindre 100 mM d'EDTA. L'augmentation de l'absorbance à 280 nm semble indiquer la présence de protéine en grande quantité (**figure 59, droite**). A titre de comparaison, l'utilisation du protocole mis au point pour INFLWT-pET-30a ne s'accompagne que d'une très légère augmentation de l'absorbance à 280 nm, en partie liée à l'augmentation de la concentration en imidazole et à la faible quantité de protéines éluées (**figure 59, gauche**).

Les bactéries cultivées en présence et en absence d'IPTG, les fractions solubles et insolubles issues de la lyse des bactéries, ainsi que les fractions non retenues et d'élution, ont ensuite été analysées par électrophorèse sur gel discontinu de polyacrylamide 12 % en conditions dénaturantes (**figure 60, gauche**). Il apparaît que la protéine est surproduite en grande quantité en présence d'IPTG et dans nos conditions de culture avec l'absence significative de fuite d'expression en absence de l'inducteur (**figure 60, gauche, pistes 1 et 2**). L'analyse des fractions solubles et insolubles nous indique également que la lyse a été efficace avec la présence d'une quantité plus importante d'INFLWT-pPROEX-HTa dans la fraction soluble que dans la fraction insoluble (**figure 60, gauche, pistes 3 et 4**). L'analyse de la fraction non retenue indique par ailleurs que la protéine est capable dans ces conditions de s'accrocher correctement à la colonne de nickel (**figure 60, gauche, piste 5**). La protéine semble présente dans les fractions d'élution avec un haut niveau de pureté (**figure 60, gauche, pistes 6 à 9**).

Les fractions d'élution ont ensuite été rassemblées puis dialysées contre un nouveau tampon dépourvu d'urée (50 mM Bis-Tris pH 7,0, 100 mM AmSO_4 , 10 mM MgSO_4 et 10 % V/V glycérol d'après (McCord *et al.* 1998)). La protéine est très stable dans ce tampon et au final, environ 10 mg de protéine par litre de culture ont ainsi été produits et purifiés avec une haute pureté. La protéine peut facilement être concentrée jusqu'à plus de 15 mg/ml par l'utilisation d'un système d'ultrafiltration avec une membrane en cellulose régénérée. Il est toutefois apparu que les membranes en Poly Ether Sulfone (PES) étaient inadaptées avec une adhésion importante de la protéine à ce support au cours de la concentration.

Une partie de la protéine purifiée a été directement congelée et stockée à -80°C dans des cryo-tubes de polypropylène. Le reste a été conservé afin de pouvoir initier des essais de clivage de l'étiquette de purification par la protéase du TEV.

Pour cela nous avons surproduit et purifié la protéase du TEV, selon un protocole déjà mis au point au sein du laboratoire et légèrement modifié, afin d'inclure une étape de lavage à forte concentration en sel (1 M NaCl) pour éliminer les contaminations par les acides nucléiques. Brièvement, une boîte LB-Agar supplémenté avec de l'ampicilline (100 $\mu\text{g/ml}$) est utilisée pour étaler le stock glycérol de bactéries BL21 DE3 transformées par le plasmide pRK793 codant la protéase du TEV. Un des clones obtenus est utilisé pour inoculer une préculture de 10 ml de milieu LB supplémenté avec de l'ampicilline (100 $\mu\text{g/ml}$) qui est mise sous agitation sur la nuit à 37°C . Le lendemain, la pré-culture est utilisée pour inoculer 1 l de milieu LB supplémenté avec de l'ampicilline (100 $\mu\text{g/ml}$) puis laissée sous agitation à 37°C jusqu'à atteindre une densité optique de 0,5 à 600 nm. De l'IPTG est ajouté (0,5 mM final) et la culture laissée à 37°C pendant 4 h. Les bactéries sont ensuite lysées par sonication dans un tampon 20 mM Tris HCl pH 7,5, 10 % V/V glycérol, 20 mM imidazole et 300 mM NaCl. La fraction soluble est injectée sur une colonne de nickel, puis un rinçage est réalisé avec le tampon de lyse modifié contenant 1 M de NaCl. L'élution de la protéine est ensuite réalisée par un gradient avec un tampon d'élution identique au tampon de lyse mais avec 300 mM d'imidazole. Le résultat de la purification indique l'obtention de la protéase avec un haut niveau de pureté (**figure 60, droite**). Les échantillons sont ensuite dialysés contre le tampon de lyse sans imidazole et du glycérol ajouté (50 % V/V final) avant de stocker les fractions concentrées à 0,75 mg/ml à -80°C dans des cryo-tubes de polypropylène.

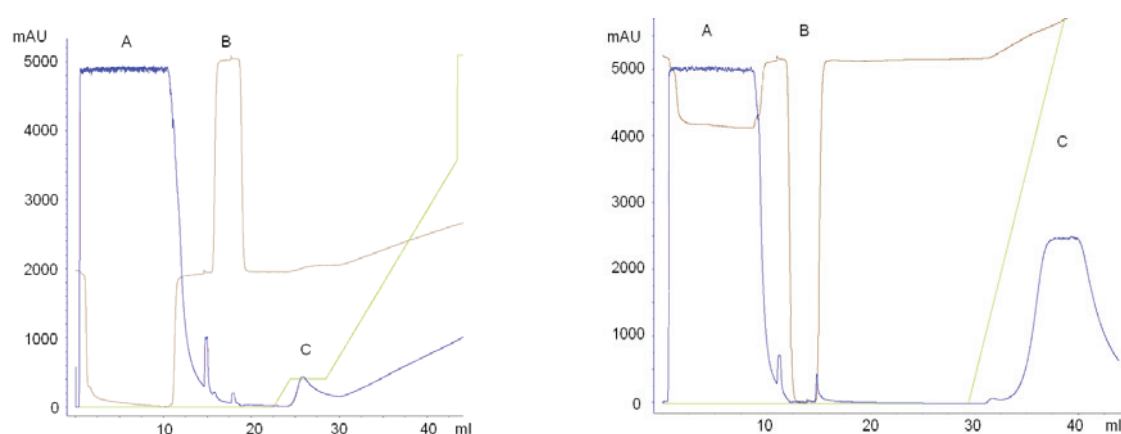


Figure 59 : Chromatogramme de la purification d'INFLWT-pPROEX-HTa par chromatographie d'affinité pour le nickel suivant le protocole utilisé pour INRAV1-pET-30a (**gauche**) et suivant le protocole mis au point pour INFLWT-pPROEX-HTa (**droite**). Courbe bleue : suivi de l'absorbance en sortie de colonne pour une longueur d'onde de 280 nm (axe des ordonnées) et en fonction du volume d'élution (axe des abscisses). Courbe marron : suivi de la conductance du tampon en sortie de colonne. Courbe verte : suivi du gradient tampon T1/tampon T3 exprimé en pourcentage V/V de tampon T3. Fraction A : fraction non retenue. Fraction B : fraction de lavage. Fraction C : fraction d'élution.

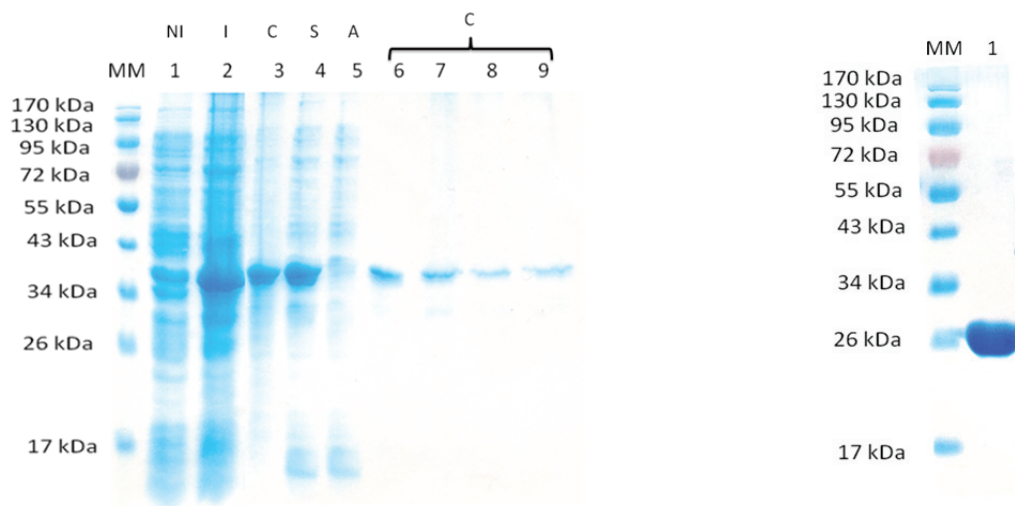


Figure 60 : (Gauche) Analyse par électrophorèse sur gel discontinu de polyacrylamide (12 %) en conditions dénaturantes, de la purification par chromatographie d'affinité pour le nickel d'INFLWT-pPROEX-HTa avec révélation par coloration au bleu de coomassie. (A) Piste MM : marqueurs de masses moléculaires connues. Piste 1 : témoin non induit. Piste 2 : témoin induit. Piste 3 : fraction insoluble (témoin induit). Piste 4 : fraction soluble (témoin induit). Piste 5 : fraction de lavage (fraction A sur le chromatogramme figure 59, droite). Piste 6 à 9 : fractions d'élution (fractions C sur le chromatogramme figure 59, droite). NI et I indiquent respectivement les fractions non induites (sans IPTG) et induites (avec IPTG). C et S indiquent respectivement la fraction insoluble (culot) et soluble (surnageant). A et C indiquent respectivement les fractions A et C du chromatogramme figure 59, droite.

(Droite) Analyse par électrophorèse sur gel discontinu de polyacrylamide (12 %) en conditions dénaturantes, du résultat de la purification de la protéase du TEV, avec révélation par coloration au bleu de coomassie. Piste MM : marqueurs de masses moléculaires connues. Piste 1 : protéase du TEV purifiée.

2.4. Clivage de l'étiquette de purification

Les conditions permettant le clivage de l'étiquette de purification d'INFLWT-pPROEX-HTa ont du être déterminées puis optimisées afin d'aboutir à un clivage total de l'étiquette. Pour cela, différents ratios protéase/protéine cible ainsi que différents temps et températures d'incubations ont été testés. Nous avons commencé les essais par l'utilisation des ratios protéase/protéine cible W/W 1/50 et 1/100 pour des temps d'incubation de une heure à sept heures et pour une température d'incubation correspondant à la température ambiante. Les fractions incubées ou non avec la protéase ont ensuite été analysées par électrophorèse sur gel discontinu de polyacrylamide 12 % en conditions dénaturantes (**figure 61**).

L'analyse du gel indique que la protéine purifiée, avant action de la protéase, est présente avec un haut niveau de pureté, mais avec quelques sous-bandes visibles qui pourraient correspondre, comme pour INFLWT-pET-30a, au résultat d'une activité de protéolyse résiduelle. L'action de la protéase du TEV est bien détectée sur gel avec l'apparition progressive d'une sous-bande correspondant à notre protéine d'intérêt dépourvue de l'étiquette de purification (**figure 61, gauche**). Le clivage ne semble pas total au bout de sept heures quelque soit le ratio utilisé et n'est pas significativement amélioré par la prolongation du temps d'incubation (données non montrées). Le changement de la température d'incubation à 4 °C réduit considérablement l'activité de clivage et ne semble pas adaptée à notre protéine (données non montrées). Nous avons donc par la suite décidé d'utiliser un ratio W/W 1/20 protéase/protéine cible, une incubation à température ambiante et une durée d'incubation sur la nuit. Dans ces conditions, le clivage a été total sans apparition d'activité de protéolyse supplémentaire (**figure 61, droite**). Ces conditions de clivage ont donc été

sélectionnées pour le clivage de l'étiquette de purification d'INFLWT-pPROEX-HTa. Un séquençage de la protéine nous a permis de confirmer le site de coupure de l'étiquette de purification.

La protéase du TEV possédant une étiquette de purification six histidines, nous avons pu réinjecter la solution issue du clivage sur une colonne de nickel, afin de récupérer notre protéine d'intérêt sans l'étiquette de purification et d'éliminer la protéase du TEV et l'étiquette de purification clivée. L'ajout d'une faible concentration d'imidazole était nécessaire pour décrocher notre protéine d'intérêt, qui présentait encore une faible affinité pour la colonne. Avant de stocker la protéine clivée à -80 °C, une dialyse finale a été réalisée afin d'enlever l'imidazole présent dans l'échantillon.

Les intégrases INFLWT-pPROEX-HTa avec et sans étiquettes ont été testées pour leurs activités de clivage et de transfert de brins par le Dr Corinne Ronfort. De façon intéressante, il est apparu que l'intégrase avec étiquette était inactive pour ces activités. Le lot sans étiquette a également été testé et présentait quant à lui des activités comparables avec INFLWT-pET-30a (données non montrées). L'étiquette de purification pourrait interférer avec la fixation des substrats de l'intégrase l'empêchant d'être active.

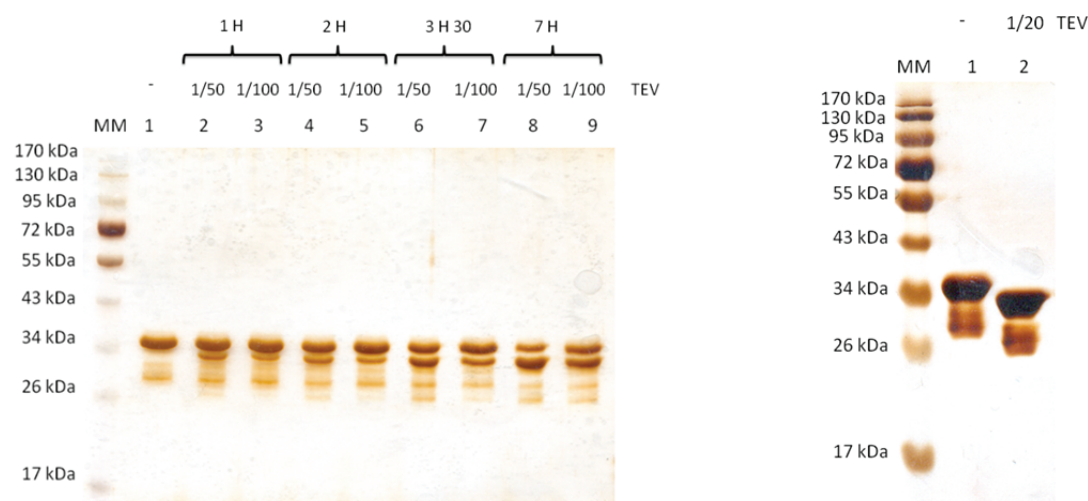


Figure 61 : Analyse par électrophorèse sur gel discontinu de polyacrylamide (12 %) en conditions dénaturantes, des échantillons d'INFLWT-pPROEX-HTa incubés en présence ou non de la protéase du TEV et avec révélation au nitrate d'argent. **(Gauche)** Piste MM : marqueurs de masses moléculaires connues. Piste 1 : fraction non incubée avec la protéase du TEV. Pistes 2, 4, 6, 8 : fractions incubées avec la protéase du TEV (ratio W/W 1/50) respectivement 1 H, 2 H, 3 H 30 et 7 H à température ambiante. Pistes 3, 5, 7, 9 : fractions incubées avec la protéase du TEV (ratio W/W 1/100) respectivement 1 H, 2 H, 3 H 30 et 7 H à température ambiante. **(Droite)** Piste MM : marqueurs de masses moléculaires connues. Piste 1 : fraction non incubée avec la protéase du TEV. Piste 2 : fraction incubée avec la protéase du TEV (ratio W/W 1/20) sur la nuit et à température ambiante.

2.5. Caractérisation de l'échantillon

2.5.1. Analyse par pontage covalent au Bis SulfoSuccinimidyl Suberate (BS3) et glutaraldéhyde

La technique du pontage covalent a été réalisée afin de caractériser les états d'oligomérisation d'INFLWT-pPROEX-HTa, afin de pouvoir comparer les résultats obtenus avec INFLWT-pET-30a. Pour cela, une analyse par électrophorèse sur gel discontinu de polyacrylamide 10 % en conditions dénaturantes, a été réalisée après avoir effectué un pontage covalent des protéines grâce à l'agent chimique BS3 ou au glutaraldéhyde. Une solution d'intégrase sans son étiquette de purification et à basse concentration (0,5 mg/ml) a été soumise au pontage par le BS3 ou au glutaraldéhyde puis déposée sur un gel de polyacrylamide (**figure 62**).

Après révélation au nitrate d'argent, on peut observer pour les agents pontant testés, la présence d'une bande présentant une masse moléculaire proche de 70 kDa et d'une bande présentant une masse moléculaire proche de 34 kDa. Ces résultats suggèrent que dans ces conditions expérimentales, l'intégrase serait majoritairement dimérique avec une forme monomérique minoritaire. Les conditions de pontage semblent optimales par l'utilisation du glutaraldéhyde 1 X avec une bande dimérique majoritaire. L'augmentation de la concentration en glutaraldéhyde (2 X) semble au contraire aboutir à l'agrégation de la protéine. Malgré la présence d'ammonium dans la solution, pouvant rentrer en compétition avec les amines des protéines pour la réaction avec les agents pontants, la réaction de clivage semble optimale dans nos conditions expérimentales. L'intégrase INFLWT-pPROEX-HTa sans son étiquette de purification semble donc bien adopter le même état oligomérique que INFLWT-pET-30a.

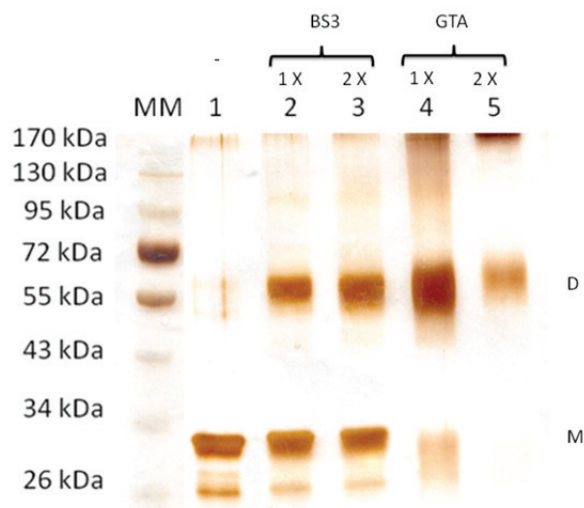


Figure 62 : Analyse de l'état d'oligomérisation de l'intégrase INFLWT-pPROEX-HTa (0,5 mg/ml) sans étiquette de purification par électrophorèse sur gel discontinu de polyacrylamide (10 %) en conditions dénaturantes et après pontage par le BS3 ou par le glutaraldéhyde avec révélation par coloration au nitrate d'argent. Piste MM : marqueurs de masses moléculaires connues. Piste 1 : contrôle sans agent pontant. Piste 2 : pontage au BS3 (1 X). Piste 3 : pontage au BS3 (2 X). Piste 4 : pontage au glutaraldéhyde (1 X). Piste 5 : pontage au glutaraldéhyde (2 X). M et D indiquent respectivement les formes monomériques et dimériques.

2.5.2. Analyse par spectroscopie à corrélation de photons

Les échantillons d'INFLWT-pPROEX-HTa et sans étiquette, ont été analysés par spectroscopie à corrélation de photons afin de vérifier la monodispersité de l'échantillon (**figure 63**).

A une concentration de 1 mg/ml, nous pouvons observer la présence d'un pic sur les profils de distribution du volume, caractérisé par un indice de polydispersité proche de 0,2. Ce résultat nous indique que la solution d'intégrase semble monodisperse.

La valeur moyenne du diamètre hydrodynamique obtenue à partir de trois mesures successives est de près de 8 nm. Cette valeur est similaire à celle obtenue pour la construction INFLWT-pET-30a et semble indiquer un état d'oligomérisation et un comportement en solution proche d'INFLWT-pET-30a.

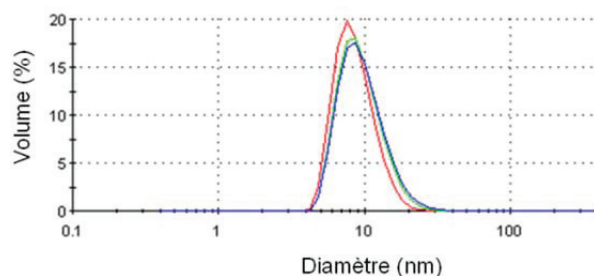


Figure 63 : Profil de distribution du volume (%) en fonction du diamètre hydrodynamique obtenu par diffusion dynamique de la lumière (superposition de trois mesures successives) pour INFLWT-pPROEX-HTa sans étiquette (1 mg/ml).

2.6. Essais de cristallisation

De nombreuses conditions de cristallisation ont été testées pour INFLWT-pPROEX-HTa sans étiquette par la technique de diffusion de vapeur et en goutte assise. Les kits commerciaux JCSG+ suite, MbClass 1 suite, MbClass 2 suite (qiagen), AmSO₄ suite et MPD suite (*Qiagen*) ont été testés et après plusieurs jours, des cristaux sont apparus dans de nombreuses conditions de cristallisation différentes. Malheureusement, l'ensemble des cristaux testés par diffraction des rayons X se sont avérés être des cristaux de sels. Il est probable que la présence de magnésium et de sulfate d'ammonium dans le tampon de la protéine a abouti à la formation de nombreux cristaux de sels dans les conditions contenant en particulier des ions phosphate. Ces cristaux ont également été obtenus par des essais réalisés sur le tampon seul en absence de protéine. Aucun cristal de protéine n'a pu être observé à ce stade de l'étude.

2.7. Conclusion

L'obtention de l'intégrase entière dans la nouvelle construction (INFLWT-pPROEX-HTa) sous une forme soluble et en grande quantité, a nécessité de nombreuses étapes d'optimisation, la protéine étant majoritairement insoluble dans des conditions optimales déterminées pour INFLWT-pET-30a. Suite à la purification d'INFLWT-pPROEX-HTa et l'obtention d'une forme sans étiquette,

nous avons voulu relancer des expériences de SAXS sur cette nouvelle construction. Cependant, au cours de ces optimisations, un modèle SAXS a alors été publié par Bojja et collaborateurs sur cette même enzyme (Bojja *et al.* 2011).

Le modèle montre des valeurs similaires à celles obtenues pour INFLWT-pET30a au niveau de la valeur du rayon de giration (Rg) et de la distance inter-atomique maximale de la particule (Dmax). En effet, les valeurs de Rg et de Dmax données par les auteurs sont respectivement de 32,8 et 109,4 Å, ces valeurs étant très proches de celles que nous avons obtenues (respectivement 33 et 110 Å). Les auteurs ont également pu placer les domaines de chaque protomère dans l'enveloppe du dimère en ajoutant des contraintes déterminées par des expériences de pontages chimiques et de spectrométrie de masse. Ils décrivent ainsi un modèle de dimère où *reaching dimer* au sein duquel les monomères interagissent de façon tête-bêche, avec les domaines catalytiques présents aux extrémités du dimère (**figure 64, milieu**). Il apparaît que l'organisation de ce dimère ressemble à l'organisation du dimère interne au sein de l'intasome de PFV (**figure 64, droite**) (Hare *et al.* 2010). L'enveloppe SAXS (calculée avec le programme GASBOR) que les auteurs ont obtenu, semble ainsi très similaire à celle que nous avons obtenue (calculée avec le programme DAMMIF).

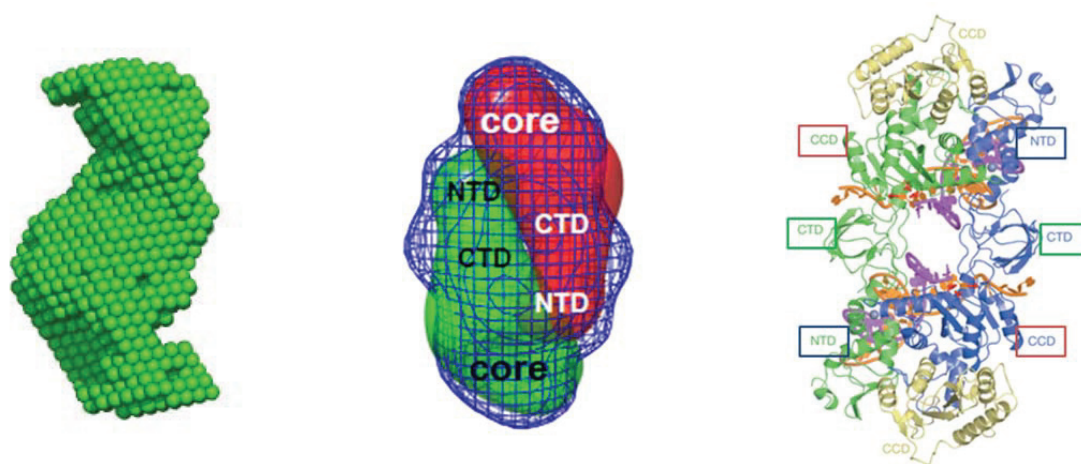


Figure 64 : (**Gauche**) Modélisation de l'enveloppe SAXS du dimère d'INFLWT-pET-30a avec le logiciel DAMMIF. (**Milieu**) Modèle SAXS du dimère de l'intégrase entière du RSV calculé avec GASBOR et proposé par Bojja et collaborateurs (Bojja *et al.* 2011). (**Droite**) Structure cristalline de l'intasome du PFV (Hare *et al.* 2010).

Suite à la publication de ces résultats, nous avons dès lors décidé de concentrer nos efforts sur l'intégrase entière du RAV-1, associée suivant la « nouvelle » interface dimérique.

3. Intégrase INFLH103C-pPROEX-HTa

3.1. Purification et caractérisation de l'échantillon

Nous avons souhaité vérifier si la protéine entière était capable, comme pour le domaine catalytique isolé, de s'assembler suivant la nouvelle interface. Pour cela, une stratégie similaire à celle utilisée pour IN-CCD-H103C, a été mise en place. Un mutant de la protéine intégrase entière du RAV-1 (mutant INFLH103C-pPROEX-HTa) a donc été surproduit en BL21 DE3 pLysS et purifié dans les mêmes conditions que celles décrites pour la protéine native.

La protéine a été correctement surproduite et purifiée et le protocole utilisé pour la protéine sauvage a également permis l'obtention d'une protéine pure sans étiquette et en quantité satisfaisante. Des expériences de pontage covalent ont révélé que la protéine mutée était toujours dimérique mais en plus faible proportion que pour l'intégrase sauvage (**figure 65, gauche**). Nous avons ensuite réalisé une analyse par électrophorèse sur gel de polyacrylamide 12 % en conditions dénaturantes et en présence ou non d'agent réducteur et après incubation ou non avec du peroxyde d'hydrogène (2 mM, trente minutes, température ambiante) (**figure 65, droite**). Il apparaît après analyse du gel (aussi bien pour la protéine native que mutée) la présence d'une forme dimérique en absence d'agent réducteur, cette forme dimérique étant plus importante pour la protéine sauvage que la protéine mutée.

Une forme trimérique minoritaire semble également apparaître en particulier pour la protéine sauvage. La présence de peroxyde d'hydrogène ne semble pas modifier les proportions de ces formes. Ces formes disparaissent en présence d'agent réducteur ce qui semble bien impliquer la formation de ponts disulfures. Le résultat obtenu n'était toutefois pas celui attendu. En effet, le contrôle négatif (protéine sauvage) semble indiquer la formation d'un dimère stabilisé par un ou des ponts disulfures. Cette forme dimérique semble moins intense pour le mutant H103C. Il apparaît donc que la protéine sauvage sous forme dimérique semble naturellement stabilisée par des ponts disulfures. La présence d'une quantité moindre de dimères après pontage covalent pour le mutant H103C pourrait ainsi expliquer l'intensité moindre de dimère stabilisé par un pont disulfure pour ce mutant. La nouvelle interface ne semble donc pas se former dans nos conditions expérimentales pour le mutant H103C mais les résultats obtenus ont souligné la présence potentielle d'un ou de plusieurs ponts disulfures inter-moléculaires sur la protéine entière sauvage.

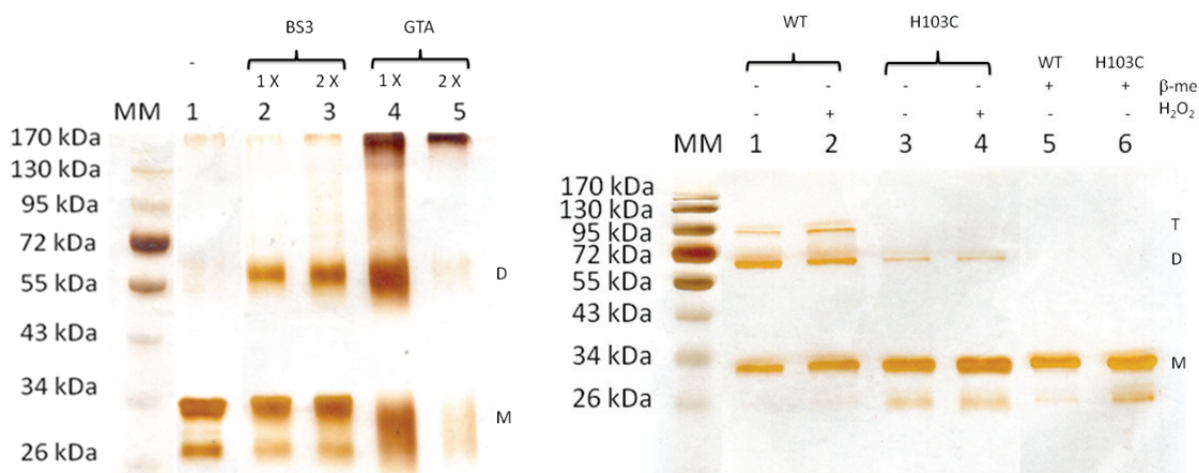


Figure 65 : (Gauche) Analyse de l'état d'oligomérisation d'INFLH103C-pPROEX-HTa (0,5 mg/ml) sans étiquette de purification par électrophorèse sur gel discontinu de polyacrylamide (10 %) en conditions dénaturantes et après pontage par le BS3 ou par le glutaraldéhyde avec révélation par coloration au nitrate d'argent. Piste MM : marqueurs de masses moléculaires connues. Piste 1 : contrôle sans agent pontant. Piste 2 : pontage au BS3 (1 X). Piste 3 : pontage au BS3 (2 X). Piste 4 : pontage au glutaraldéhyde (1 X). Piste 5 : pontage au glutaraldéhyde (2 X). M et D indiquent respectivement les formes monomériques et dimériques.

(Droite) Analyse par électrophorèse sur gel discontinu de polyacrylamide (12 %) en conditions dénaturantes d'INFL-pPROEX-HTa sauvage et H103C. Piste MM : marqueurs de masses moléculaires connues. Piste 1 et 2: protéine sauvage purifiée et analysée en conditions non réductrices respectivement sans et avec peroxyde d'hydrogène. Pistes 3 et 4: protéine H103C purifiée et analysée en conditions non réductrices respectivement sans et avec peroxyde d'hydrogène. Pistes 5: protéine sauvage purifiée et analysée en conditions réductrices sans peroxyde d'hydrogène. Pistes 6: protéine H103C purifiée et analysée en conditions réductrices sans peroxyde d'hydrogène. M, D et T indiquent respectivement les formes monomériques, dimériques et trimériques.

3.2. Conclusion

La mise au point du protocole de suproduction et de purification d'INFLWT-pPROEX-HTa a permis l'obtention de la protéine en grande quantité, sous une forme soluble et sans son étiquette de purification. Ce protocole a pu être appliqué pour le mutant INFLH103C-pPROEX-HTa permettant l'obtention de la protéine avec un bon niveau de pureté et en quantité suffisante pour permettre de lancer des études biochimiques et structurales (en particulier SAXS).

Nous avons dès lors tenté de mettre en évidence la formation de la nouvelle interface dimérique par l'intermédiaire d'un pont disulfure inter-moléculaire C103-C103 sur la protéine entière en utilisant le même protocole que celui développé pour IN-CCD-H103C. La protéine a donc été incubée avec du peroxyde d'hydrogène dans les conditions optimales déterminées pour IN-CCD-H103C. Nous avons ensuite vérifié la présence d'un pont disulfure sur INFLH103C-pPROEX-HTa par électrophorèse sur gel de polyacrylamide en conditions dénaturantes et en présence ou non d'agent réducteur. Cette expérience a également été réalisée sur la protéine sauvage INFLWT-pPROEX-HTa. Les résultats obtenus indiquent la présence d'une forme dimérique stabilisée par un pont disulfure sur INFLH103C-pPROEX-HTa mais également sur la protéine sauvage INFLWT-pPROEX-HTa. Différentes conditions (de pH, de concentration en peroxyde d'hydrogène, de concentration protéique) ont été testées mais n'ont pas abouti à des modifications significatives de ces résultats.

La nouvelle interface ne semble donc pas se former dans nos conditions expérimentales pour le mutant H103C mais les résultats obtenus ont souligné la présence potentielle d'un ou plusieurs ponts disulfures sur la protéine entière sauvage.

CONCLUSION ET PERSPECTIVES

Au cours de cette thèse, je me suis intéressé à la protéine intégrase du virus RAV-1 (ou IN RAV-1) appartenant au genre *Alpharétrovirus* des *Retroviridae*. Cette intégrase est étudiée depuis 2002 à l'IBCP, au sein d'une collaboration entre les équipes de « Biocristallographie et Biologie Structurale des Cibles Thérapeutiques » et « Rétrovirus, Intégration Virale et Vaccins » du pôle scientifique de Lyon-Gerland.

Etude du domaine catalytique central (CCD) d'IN RAV-1

Un « nouvel » assemblage dimérique

La première partie de mon travail de thèse a porté sur l'étude du nouvel assemblage du CCD isolé d'IN RAV-1 et s'est inscrite dans la continuité du travail de thèse du Dr A. Ballandras. J'ai mis en évidence que ce « nouvel » assemblage pouvait se former en solution, grâce à l'utilisation d'un mutant « cystéine » (mutant H103C), permettant de stabiliser la « nouvelle » interface dimérique par un pont disulfure. Cette stratégie appelée « *cysteine-directed mutagenesis* », a déjà été employée sur d'autres protéines, pour mettre en évidence les interfaces d'oligomérisation impliquées dans différents assemblages macromoléculaires (Jilaveanu & Oliver 2006; Markovic & Dutzler 2007; Lee *et al.* 2008).

La « nouvelle » interface dimérique est significativement différente de l'interface « canonique » et nécessite un déplacement d'une dizaine d'angströms de chaque monomère. Il faut noter que le CCD des IN des rétrovirus de l'espèce aviaire (comme RAV-1 et RSV) sont monomériques en solution pour des concentrations de l'ordre du mg/ml (manuscrit de thèse d'A. Ballandras et (Bujacz *et al.* 1995)), alors que le CCD d'IN VIH-1 est majoritairement dimérique à cette concentration. Nous pensons donc que l'association dimérique du CCD des IN des rétrovirus de l'espèce aviaire, sous leur forme « nouvelle » ou « canonique », se produit au moment de la cristallisation.

Les cristaux du « nouvel » arrangement dimérique ont été obtenus en présence d'ions Zn^{2+} dans un tampon MES à pH 6,0. L'analyse de la structure cristalline montre que l'assemblage est stabilisé en son centre par un ion Zn^{2+} , qui est coordonné par deux histidines symétriques (H103). Le résidu H103, qui est très conservé au sein des IN d'alpharétrovirus, semble ainsi jouer un rôle crucial dans l'assemblage de la nouvelle forme dimérique par sa capacité à coordonner le zinc. Un ion zinc avait également été observé à proximité du résidu H103 sur la forme dimérique « canonique » du CCD d'IN RSV ainsi que dans le site actif de l'enzyme (Bujacz *et al.* 1997; Lubkowski *et al.* 1998). De façon intéressante, on retrouve dans la structure du « nouvel » assemblage dimérique, la présence d'un ion Zn^{2+} dans le site actif de chaque monomère. Le zinc est un élément connu comme étant nécessaire aux intégrases puisqu'il intervient dans le repliement du domaine N-terminal (Jaskolski *et al.* 2009). De plus, bien que le magnésium soit considéré comme le cofacteur naturel de l'activité catalytique de l'IN, le zinc peut également, dans une moins mesure, contribuer à la réaction d'intégration en qualité de cofacteur (Bujacz *et al.* 1997). L'observation de ces ions Zn^{2+} dans la « nouvelle » interface et le site actif d'IN RAV-1 pourrait donc avoir une cohérence biologique.

Les dernières données structurales sur l'IN entière du PFV, qui est active *in crystallo*, montrent la nécessité d'une interface de dimérisation « canonique » des CCD au sein du tétramère, pour les étapes de clivage et de transfert de brins (Maertens *et al.* 2010). Nous pensons donc que le « nouvel » arrangement dimérique, devrait entraîner une réorganisation des autres domaines (CTD et NTD) sur la protéine entière, la rendant inactive pour l'étape d'intégration. Cette « nouvelle » structure quaternaire pourrait cependant intervenir dans une autre étape du cycle réplcatif rétroviral, l'IN ayant été décrite comme possédant des fonctions pléiotropiques (Masuda 2011).

La présence d'un sillon fortement basique à l'interface de dimérisation et une étude de *docking* suggèrent ainsi un rôle de ce « nouvel » assemblage dans la fixation d'acides nucléiques simples brins, ARN ou ADN (Ballandras *et al.* 2011). Pour étayer cette hypothèse, j'ai mené différents essais de trempages et de co-cristallisations avec différents nucléotides et désoxyribonucléotides. Toutefois, les cartes de densités électroniques obtenues après enregistrements des données rayons X et affinements cristallographiques, n'ont pas permis d'observer d'acides nucléiques liés à la « nouvelle » interface. Nous avons cependant pu observer suite à ces trempages, la présence d'un ou deux phosphates coordonnés à l'ion Zn^{2+} central reliant les résidus H103 de chaque monomère. Nous ne savons pas si ces observations présentent une signification biologique. Comme les essais de trempages et de co-cristallisation ont été réalisés à un pH acide (pH 6,0), il est envisageable que ces phosphates proviennent de l'hydrolyse des nucléotides sous l'effet d'une hydrolyse acide possible à ce pH. Des études supplémentaires sont nécessaires pour déterminer le rôle potentiel de la fixation de ce ou de ces phosphates à la « nouvelle » interface de dimérisation.

Vers des inhibiteurs allostériques

Le « nouvel » assemblage a été cristallisé en présence de tampon MES et cette molécule a été observée dans une crevasse située à l'interface de dimérisation (Ballandras *et al.* 2011). L'hypothèse d'une forme inactive de l'enzyme sous cette conformation a suscité l'idée d'utiliser le MES comme tête de série, pour le développement d'un nouveau type d'inhibiteurs allostériques. En effet, des composés développés à partir du MES et présentant une très forte affinité pour l'interface, pourraient être utilisés pour stabiliser ce nouvel assemblage dans une conformation inactive pour l'enzyme et ainsi perturber le cycle réplcatif rétroviral.

Nous avons utilisé le mutant disulfure H103C pour mesurer l'interaction du MES avec le « nouvel » assemblage dimérique, grâce à des techniques d'ITC, d'atténuation de fluorescence et de spectroscopie de fluorescence différentielle à balayage thermique, que j'ai mises au point. Ces techniques n'ont cependant pas permis de mettre en évidence d'interaction protéine-ligand jusqu'à une concentration de 100 mM de composé, qui est équivalente à celle retrouvée dans la solution de cristallisation.

De façon intéressante, de nombreux auteurs ont observé la fixation du composé MES dans le site actif des enzymes ou à leurs surfaces. Un exemple est celui de la structure cristalline de la β -lactamase de *Bacteroides fragilis*, avec une molécule de tampon MES liée dans le site actif. Les auteurs ont réussi à déterminer un K_i de 25 mM pour ce composé et suggèrent d'utiliser le MES comme base pour le développement de nouveaux inhibiteurs compétitifs (Fitzgerald *et al.* 1998). Une autre étude réalisée par RMN a mis en évidence la possibilité d'interférence du tampon avec la protéine, comme l'interaction du MES avec une protéine du foie liant les acides gras chez l'Homme (hLFABP pour *Human Liver Fatty Acid Binding Protein*). Une titration de cette enzyme avec ce composé par RMN, a indiqué un phénomène d'interaction ne s'accompagnant pas d'une saturation pour une concentration maximale de 50 mM (Long & Yang 2009). Il est donc envisageable que l'interaction du MES avec notre « nouvel » assemblage ne puisse pas être détectée à la concentration maximale testée de 100 mM et ce malgré son observation *in crystallo*

Comme décrit précédemment, le MES est observé dans le site actif de structures cristallines, où il peut mimer la structure chimique d'un substrat ou d'un cofacteur. En particulier, différents exemples existent dans la littérature où le groupement sulfonate du MES mime la position d'un phosphate ou d'un sulfate « biologique » (Toney *et al.* 1993; Knochel *et al.* 1996). Il est donc également envisageable que le MES puisse mimer, dans la « nouvelle » interface, la position d'une molécule biologique constituée d'un ou plusieurs groupements phosphates ou sulfates.

Sur le même modèle, l'interaction d'une molécule de saccharose issue de la solution de cristallisation avec l'interface dimérique « canonique » du CCD de l'IN du VIH-1 a été rapportée (Wielens *et al.* 2010). Les auteurs ont postulé que cette molécule pourrait affecter les activités de l'IN et être utilisée pour le développement de nouveaux inhibiteurs allostériques. Ils n'ont toutefois pas pu mettre en évidence *in vitro* de phénomène d'interaction du composé avec cette IN. Le saccharose n'a par ailleurs révélé aucun effet inhibiteur sur le cycle réplcatif rétroviral du VIH-1 (Wielens *et al.* 2010). De la même façon, les effets du MES, ainsi que de composés analogues synthétisés par nos collaborateurs, ont été évalués sur les activités catalytiques de l'IN et sur le cycle réplcatif rétroviral du RAV-1 par l'équipe du Dr C. Ronfort mais aucun effet inhibiteur n'a été observé. Le développement de composés de plus forte affinité semble donc nécessaire.

De nouvelles études de *docking* mises en place par nos collaborateurs chimistes ont montré que le Raltégravir était un bon candidat pour se lier à la place du MES dans la « nouvelle » interface. De nouveaux composés ont été synthétisés à partir du Raltégravir, mais en éliminant la fonction dicétone, afin que ceux-ci ne puissent pas agir au niveau du site actif de l'enzyme en coordonnant les cations divalents. Des expériences *in vitro* (fluorimétrie différentielle à balayage thermique, gel natif) et *in cristallo* (trempages et co-cristallisations) ont été réalisées, mais aucune interaction n'a pu être observée entre ces composés et le « nouvel » assemblage. Par ailleurs, nous n'avons pas pu observer de stabilisation de la forme dimérique du CCD sauvage d'IN-RAV1 en présence de ces composés. Le développement de nouvelles molécules de taille intermédiaire entre le MES et le Raltégravir est en cours. Leur interaction avec la « nouvelle » interface pourra être validée par les méthodes mises au point au cours de ce travail de thèse.

Par ailleurs l'histidine 103, qui est très conservée pour les alpharétrovirus ne semble pas retrouvée chez les autres rétrovirus en particulier chez le VIH-1 où elle est remplacée par une phénylalanine. Une modélisation de l'IN du VIH-1 suivant le « nouvel » assemblage montre cependant la possibilité d'une telle association. Par ailleurs, le sillon basique est également retrouvé sur le modèle et les résidus du sillon semblent conservés. Nous pouvons ainsi imaginer qu'un tel assemblage sur le VIH-1 pourrait se produire au cours du cycle réplcatif rétroviral mais nécessiter la présence d'un partenaire cellulaire pour être stabilisé.

Des expériences de « *cysteine-directed mutagenesis* » sont envisagées pour vérifier cette hypothèse. L'existence de cet assemblage sur l'IN du VIH-1 pourrait ainsi présenter de nouvelles perspectives d'inhibition de cette enzyme par des inhibiteurs allostériques qui pourraient compléter les inhibiteurs actuellement utilisés en multithérapie anti-VIH.

Etude de l'IN entière du RAV-1

Une autre partie de mes recherches a été consacrée à l'étude de l'IN entière du RAV-1, pour laquelle des données structurales n'étaient disponibles que pour des fragments de la protéine.

Un protocole de surproduction et de purification a été mis au point permettant l'obtention d'un bon rendement de purification. Une caractérisation biophysique de la protéine a permis de déterminer que l'échantillon était monodisperse et que la protéine était majoritairement dimérique en solution à basse concentration (en dessous de 3 mg/ml). Ces résultats sont cohérents avec ceux disponibles dans la littérature où l'IN du RSV est décrite comme étant majoritairement dimérique en solution en présence d'une forte concentration de sel (500 mM NaCl) (Jones *et al.* 1992; Coleman *et al.* 1999). Des essais de cristallisation ont été initiés, mais sans succès. Toutefois, des données structurales à basse résolution ont été obtenues par SAXS et ont permis de calculer une première enveloppe de la protéine (en collaboration avec le Pr. P. Vachette, IBBMC Paris-Orsay).

Afin d'améliorer la qualité de ce modèle, une nouvelle construction a été réalisée afin d'éliminer les étiquettes de purification. Cependant, au cours de l'optimisation de la surproduction et de la purification de l'IN dans cette nouvelle construction, l'équipe de A. M. Skalka a publié un modèle SAXS sur cette même enzyme (Bojja *et al.* 2011). Le modèle décrit par ces auteurs est très proche du notre avec un rayon de giration et une distance inter-particulaire maximale très similaires. Leurs résultats indiquent que l'IN est majoritairement sous forme dimérique et ils décrivent un « *reaching dimer* » où les monomères auraient une organisation tête-bêche. Au sein de ce dimère, les domaines catalytiques n'interagissent pas ensemble, à l'inverse de ce qui est observé dans l'ensemble des structures d'INs tronquées connues jusqu'à présent.

Récemment, une autre équipe a obtenu des cristaux sur l'IN entière du RSV (Shi *et al.* 2013). Cependant, les structures résolues ne permettent pas de déterminer la position du NTD, qui n'est pas observé dans les cartes de densité électronique. Ces structures, qui présentent une IN sous forme dimérique, sont par ailleurs identiques aux structures de l'IN du RSV tronquée (CCD+CTD) et présentent une interface de dimérisation impliquant les domaines catalytiques. Ce dernier résultat n'est donc pas en adéquation avec le modèle SAXS précédemment obtenu.

De nombreuses interrogations subsistent donc sur l'organisation structurale d'IN RAV-1 ainsi que sur l'importance de l'interface de dimérisation CCD/CCD dans les assemblages des IN. De façon intéressante, une étude SAXS récente (Bojja *et al.* 2013) indique une possible flexibilité de l'interface de dimérisation de l'IN du VIH-1 qui pourrait adopter deux conformations ; une impliquant une interface de dimérisation CCD/CCD, l'autre n'impliquant pas les domaines CCD, comme dans le cas du modèle précédemment décrit de « *reaching dimer* » (selon la présence ou non de zinc dans la solution) (Bojja *et al.* 2013). Ces données pourraient indiquer une flexibilité dans les interfaces d'oligomérisation de la protéine avec de possibles implications biologiques. Une telle flexibilité avait par ailleurs déjà été observée pour le domaine N-terminal de l'IN du VIH-1 (Wang *et al.* 2001).

Etude de l'IN entière mutée H103C

Nous avons souhaité vérifier si la nouvelle interface de dimérisation pouvait se former sur l'IN entière. Pour cela une stratégie basée sur l'utilisation du mutant H103C a également été utilisée. Les résultats obtenus sur l'IN entière de RAV-1 n'étaient cependant pas ceux attendus. En effet, il est apparu que la protéine sauvage présentait des ponts disulfures naturels là où aucun pont disulfure C103-C103 n'a pu être observé sur le mutant H103C.

L'IN entière sauvage présente plusieurs cystéines libres qui pourraient participer à la formation de ces ponts disulfures. De façon intéressante, des résultats similaires avaient déjà été observés pour l'IN du VIH-1 mais les auteurs n'avaient pas pu en tirer de signification biologique (Bischerour *et al.* 2003). En effet, la mutation de la cystéine impliquée (C280) n'a pas modifié les activités de l'enzyme et n'a pas eu d'effet significatif sur le cycle de réplication du virus. Les auteurs ont toutefois suspecté un rôle potentiel de ces ponts disulfure dans la biosynthèse de la protéine et en particulier pour la formation d'intermédiaires de repliement. Il serait par ailleurs intéressant d'avoir l'identité des cystéines impliquées afin de mieux localiser les positions des domaines au sein des oligomères de l'IN.

L'absence de la nouvelle forme dimérique sur la protéine entière pourrait s'expliquer. Il a été rapporté sur le modèle SAXS de l'IN du RSV que les interfaces CCD/CCD n'étaient pas présentes sur le dimère mais étaient présentes sur le tétramère présent en faible proportion (Bojja *et al.* 2011). Il est possible que la mise en évidence de la nouvelle interface nécessite également une forme tétramérique de l'enzyme. Des études précédentes ont montré qu'en dessus de 500 mM de sel (NaCl), la forme dimérique était majoritaire mais que la forme tétramérique devenait prédominante lors du

retour des concentrations en sels à des valeurs plus physiologiques (Coleman *et al.* 1999). Des essais pourraient être réalisés à différentes concentrations en sels pour vérifier la formation du nouvel assemblage. On ne peut également pas écarter le rôle possible d'un partenaire cellulaire ou d'un acide nucléique pour stabiliser cette nouvelle conformation.

PUBLICATION 1 :

**A CRYSTAL STRUCTURE OF THE CATALYTIC CORE DOMAIN OF AN AVIAN
SARCOMA AND LEUKEMIA VIRUS INTEGRASE SUGGESTS AN ALTERNATE
DIMERIC ASSEMBLY**

Allison Ballandras, Karen Moreau, Xavier Robert, Marie-Pierre Confort, **Romain Merceron**,
Richard Haser, Corinne Ronfort, Patrice Gouet

PLOS one, 2011

A Crystal Structure of the Catalytic Core Domain of an Avian Sarcoma and Leukemia Virus Integrase Suggests an Alternate Dimeric Assembly

Allison Ballandras¹, Karen Moreau², Xavier Robert¹, Marie-Pierre Confort², Romain Merceron¹, Richard Haser¹, Corinne Ronfort^{2*}, Patrice Gouet^{1*}

¹ Biocristallographie et Biologie Structurale des Cibles Thérapeutiques, Institut de Biologie et Chimie des Protéines, UMR 5086 BMSSI-Centre National de la Recherche Scientifique/Université de Lyon, Lyon, France, ² Laboratoire "Rétrovirus et Pathologie Comparée", UMR 754-Institut National de la Recherche Agronomique/Université de Lyon, École Nationale Vétérinaire de Lyon, Lyon, France

Abstract

Integrase (IN) is an important therapeutic target in the search for anti-Human Immunodeficiency Virus (HIV) inhibitors. This enzyme is composed of three domains and is hard to crystallize in its full form. First structural results on IN were obtained on the catalytic core domain (CCD) of the avian Rous and Sarcoma Virus strain Schmidt-Ruppin A (RSV-A) and on the CCD of HIV-1 IN. A ribonuclease-H like motif was revealed as well as a dimeric interface stabilized by two pairs of α -helices ($\alpha 1/\alpha 5$, $\alpha 5/\alpha 1$). These structural features have been validated in other structures of IN CCDs. We have determined the crystal structure of the Rous-associated virus type-1 (RAV-1) IN CCD to 1.8 Å resolution. RAV-1 IN shows a standard activity for integration and its CCD differs in sequence from that of RSV-A by a single accessible residue in position 182 (substitution A182T). Surprisingly, the CCD of RAV-1 IN associates itself with an unexpected dimeric interface characterized by three pairs of α -helices ($\alpha 3/\alpha 5$, $\alpha 1/\alpha 1$, $\alpha 5/\alpha 3$). A182 is not involved in this novel interface, which results from a rigid body rearrangement of the protein at its $\alpha 1$, $\alpha 3$, $\alpha 5$ surface. A new basic groove that is suitable for single-stranded nucleic acid binding is observed at the surface of the dimer. We have subsequently determined the structure of the mutant A182T of RAV-1 IN CCD and obtained a RSV-A IN CCD-like structure with two pairs of buried α -helices at the interface. Our results suggest that the CCD of avian INs can dimerize in more than one state. Such flexibility can further explain the multifunctionality of retroviral INs, which beside integration of dsDNA are implicated in different steps of the retroviral cycle in presence of viral ssRNA.

Citation: Ballandras A, Moreau K, Robert X, Confort M-P, Merceron R, et al. (2011) A Crystal Structure of the Catalytic Core Domain of an Avian Sarcoma and Leukemia Virus Integrase Suggests an Alternate Dimeric Assembly. PLoS ONE 6(8): e23032. doi:10.1371/journal.pone.0023032

Editor: Jianming Qiu, University of Kansas Medical Center, United States of America

Received: March 9, 2011; **Accepted:** July 7, 2011; **Published:** August 9, 2011

Copyright: © 2011 Ballandras et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: AB is supported by a Ph. D. grant from the Cluster 10 Infectiology of the Rhône-Alpes Region. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: ronfort@univ-lyon1.fr (CR); p.gouet@ibcp.fr (PG)

Introduction

During the replicative cycle of retroviruses, the retrotranscribed viral DNA is integrated into the host chromosome by the viral integrase protein (IN) [1]. The integration reaction is essential for the viral life cycle; therefore, IN is a key target for antiretroviral drug design [2–5]. Retroviral integration proceeds in three steps, two of which are catalyzed by IN. First, during the 3' processing, the two 3' terminal nucleotides of each viral DNA end are removed to generate CA-3'-OH ends with a two-base 5' overhang. Then, during the strand transfer, the recessed 3'-OH viral ends attack the phosphodiester bonds of the cellular DNA at cleavage sites separated by four to six base pairs (depending on the virus) and the viral DNA is joined to the host DNA. Finally, gap filling and DNA ligation are performed, probably by cellular enzymes [6–8].

Retroviral IN consists of three domains: the zinc-binding N-terminal domain (NTD), the catalytic core domain (CCD) and the C-terminal domain (CTD). The IN proteins of the Avian Sarcoma and Leukemia Viruses (ASLV) and Human Immunodeficiency Virus (HIV) are approximately 280 amino acids long. The NTD

binds viral DNA [9] and target DNA [10,11] and promotes IN oligomerization [12]. The NTD is required for 3' processing and strand transfer *in vitro*. The central CCD contains an invariant D,D(35)E motif, which forms a catalytic triad with two sites that can coordinate various divalent cations (Mg(II), Mn(II), Zn(II), Cd(II), Ca(II)) [13–16] although Mg(II) is the likely metal cofactor *in vivo*. The CCD alone is sufficient to perform an *in vitro* reaction termed disintegration, which is the reverse of the strand transfer reaction [17]. This domain is the most conserved domain across retroviral INs (>20% sequence identity). It belongs to the ribonuclease H-like superfamily [18,19] and consists of a five-stranded mixed β -sheet flanked by α -helices. It has always been solved as a dimer in partial or entire IN structures from lentivirus (HIV-1, HIV-2, Simian Immunodeficiency Virus (SIV), Maedi-Visna Virus (MVV, Bovine Immunodeficiency Virus (BIV)), alpharetrovirus (Rous Sarcoma Virus (RSV)) and spumavirus (Prototype Foamy Virus (PFV)) with an intermolecular interface that always involves two pairs of facing α -helices [20]. The CTD is known to bind both viral DNA and target DNA [21] and is also involved in oligomerization [22].

The three domains are connected by flexible loops, making the full-length enzyme difficult to crystallize. Hence, the structure of IN was first investigated in fragments, such as the two-domains HIV-1 IN fragment [23,24] and the two-domains RSV IN fragment [25]. Recently, 3D models of negatively stained full-length HIV-1 IN, alone or complexed with the cellular cofactor LEDGF/p75 and either viral or cellular DNA, were proposed by electron microscopy [26]. The crystal structure of the full-length IN from PFV complexed with viral [16] and cell DNA [27] was determined soon thereafter. The EM and crystal structures confirm that two IN dimers are necessary for concerted integration. In each dimer, only one CCD active site binds viral DNA and performs the 3' processing and strand transfer reactions. The two remaining CCD active sites of the tetramer lie far from the bound DNA ends and have no apparent role. Taken together, these structures further suggest that the NTD and the CTD can move during integration, and their positions diverge with respect to the CCD.

Rous-associated virus type 1 (RAV-1) is a replication-competent alpharetrovirus, member of ASLV subgroup A; the INs of this retrovirus genus are a good model for HIV IN [28]. Herein, we have determined the crystal structure of the CCD of RAV-1 IN to 1.8 Å resolution. The resulting structure exhibits an unexpected new dimeric arrangement with potential biological implications. Our experimental data also explain how crystallization conditions, as well as the single amino-acid substitution A182T between the RAV-1 IN CCD and the well-studied RSV-A (strain Schmidt-Ruppin A) IN CCD [13,14], can favor either dimeric form during crystal growth. We further show by docking calculations that this novel dimeric form could accommodate a single-stranded nucleic acid.

Results

Structure determination and refinement

The CCD of RAV-1 IN, consisting of residues 53–199, was expressed in *Escherichia coli* and purified as described in the 'Materials and Methods' section. This fragment differs by a single residue from the CCD of RSV-A IN (A182T substitution), for which numerous crystal structures have been solved [13,14,29–31]. Crystallization conditions similar to those published for the CCD of RSV-A IN that is, citrate buffer at an acidic pH and HEPES buffer at an alkaline pH [31], were tried but this approach proved unsuccessful. Hence, a broad screening of conditions was performed. Crystals were obtained in the presence of Zn(II) and MES at pH 6.0. They belonged to the hexagonal space group $P6_1$ and contained two molecules in the asymmetric unit. Synchrotron data were collected to 1.8 Å resolution near the Zn-K absorption edge. The phase problem was solved by molecular replacement using the monomer of RSV-A IN CCD (PDB entry 1VSD) structure as the search model. After a few cycles of crystallographic refinement alternated with manual rebuilding, the final crystal structure was obtained with an R_{factor} value of 19.3% (R_{free} 22.8%). The structure contains 271 amino acids in two monomers termed A and B, 186 water molecules, three Zn(II) ions and one MES molecule (Figure 1A, left). The N-terminal ends 53–57 of the two monomers, as well as the loops formed by 145A–152A (monomer A) and 145B–149B (monomer B) were not observed in electron density maps and are not included in the model. These loops are often disordered in retroviral INs [32]. The three coordinated Zn(II) correspond to the highest peaks in the calculated anomalous difference Fourier map (29σ to 33σ).

As a control, the A182T mutant of RAV-1 IN CCD (termed RAV-1 IN CCD_{A182T}), corresponding to the RSV-A IN CCD

sequence, was purified and the crystallization conditions for both RAV-1 IN CCD and RSV-A IN CCD were tested; that is, a MES buffer at pH 6.0 and a citrate buffer at pH 6.2, respectively. Microcrystals were observed with the first set of conditions, but were too small to give measurable Bragg peaks. Large crystals were obtained with the second set of conditions. Further, the huge crystals of RAV-1 IN CCD_{A182T} were isomorphous to those of RSV-A IN CCD obtained under the same conditions. They belonged to space group $P4_32_12$ with one molecule in the asymmetric unit. Synchrotron data for these crystals were collected to 1.55 Å resolutions. The phase problem was solved by a simple rigid-body refinement followed by restrained refinement using the structure of RSV-A IN CCD as the starting model. The refined structure of RAV-1 IN CCD_{A182T} contains 137 residues, 1 citrate molecule and 122 water molecules (Figure 1A, right). The 145–152 loop is disordered and is not observed in the electron density map, as in RSV-A IN CCD.

Overall structure of RAV-1 IN CCD

The 1.8 Å crystal structure of RAV-1 IN CCD consists of two identical polypeptide chains, termed A and B. The two monomers can be superimposed with an r.m.s. deviation of 0.4 Å on 132 C α pairs after a 180° rotation. The main differences between the two C α traces are due to crystal contacts. The differences arise at residues 174–176, located in a turn between helices $\alpha 4$ and $\alpha 5$ (0.8–1.1 Å between C α pairs), and at residues 198–199 at the C-terminal end (3–8 Å between C α pairs). In the latter case, the short C-terminal loop following helix $\alpha 5$ folds back toward the protein core to cap a MES molecule in molecule A (Figure 2A), whereas it protrudes into the solvent in molecule B. The non-crystallographic A/B homodimer is compact and approximates a globular ellipsoid with dimensions of 55×40×40 Å (Figure 1A). Thus, the tertiary structure of RAV-1 IN CCD is nearly identical to that of RSV-A IN CCD, equivalent to that of RAV-1 IN CCD_{A182T} (r.m.s. deviation of 0.4 Å on 132 C α pairs) and respects the ribonuclease H fold (Figure 1A). The A182T substitution, which also occurs between the Schmidt-Ruppin strains B and A of RSV IN, does not affect the tertiary structure of avian INs.

The active site of RAV-1 IN CCD

The invariant acidic residues of the catalytic triad (Asp64, Asp121 and Glu157 in RAV-1 IN) were accurately orientated in the electron density map. They form a pocket at the surface of the RAV-1 IN CCD monomer, which is located 15 Å away from the A/B dimeric interface. The carboxylate group of Asp64 is situated at the centre of the triad and interacts with Asp121 through a Zn(II) ion from the crystallization solution. This ion occupies the divalent cation-binding site termed site I in retroviral INs. It also coordinates two water molecules and the imidazole group of His198, which belongs to a symmetry-related molecule (Figure 2B). Such a penta-coordinated metal ion has never been observed in the active site of INs before. The side chain of Glu157 points freely towards the solvent, as observed in the structure of RSV-A IN CCD in complex with one Mg(II) coordinated at site I (PDB entry 1VSD). This residue rotates when accommodating the second catalytic Zn(II) in site II [14]. The present active site of RAV-1 IN with its coordinated histidine is very similar to that of influenza virus polymerase, which was solved recently [33]. In this structure, acidic and basic catalytic residues coordinate two divalent cations, which are very likely to be responsible for the endonuclease activity of the whole protein. No biological interpretation could be deduced from this structural resemblance though. A further structural comparison was performed, with the active site of full-length PFV IN in complex with Mg(II) and raltegravir, an

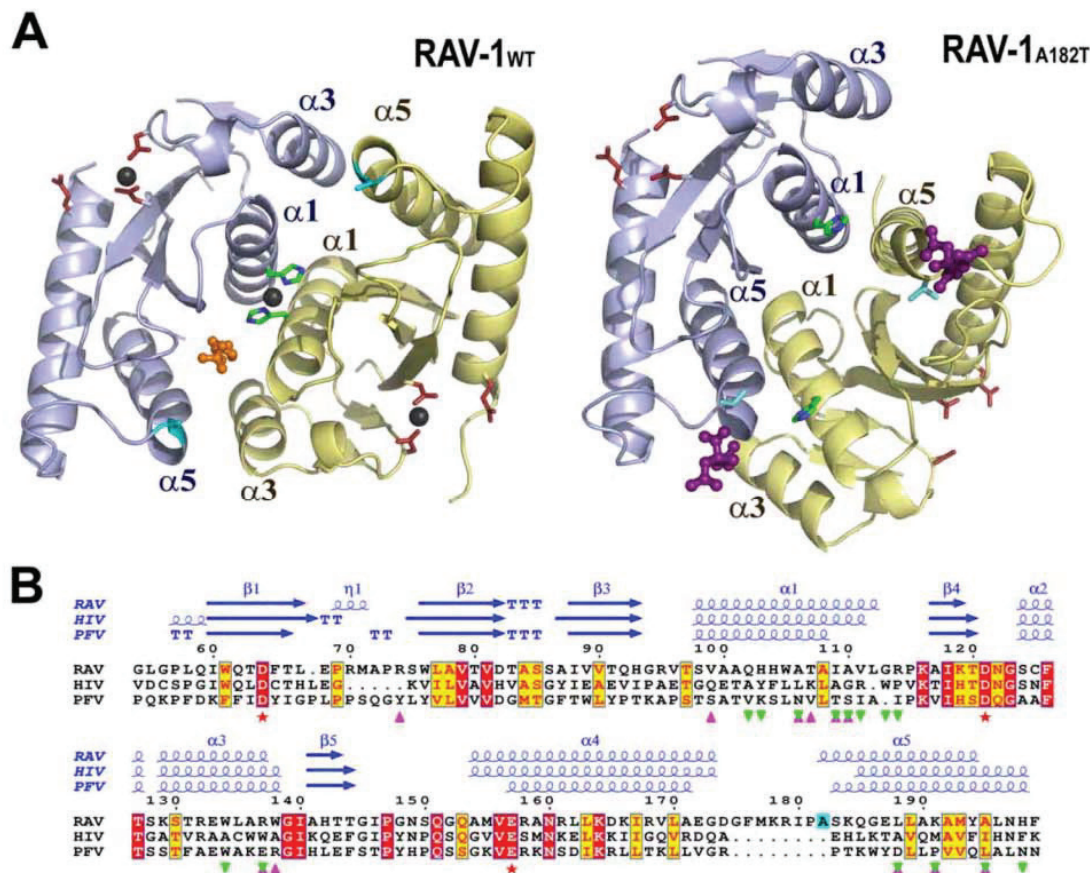


Figure 1. Primary, secondary and tertiary structure representations of RAV-1 IN CCD and RAV-1 IN CCD_{A182T} homodimers. (A) Ribbon representation of RAV-1 IN CCD (left) and RAV-1 IN CCD_{A182T} (right) with monomer A in blue and monomer B in yellow. The three catalytic residues (D64, D121 and E157) are represented by red sticks, H103 by green sticks and residue 182 (Ala in RAV-1_{WT}, Thr in RAV-1_{A182T}) by cyan sticks. The MES molecule located at the interface of RAV-1 is shown in orange, while zinc ions are represented by dark spheres. Citrate molecules are colored purple in RAV-1_{A182T}. (B) Multiple sequence alignment of RAV-1, HIV-1 and PFV IN CCDs. Secondary structure elements are indicated above the alignment. Identical and conserved residues according to physicochemical criteria are highlighted in red and yellow boxes, respectively. Pink and green triangles indicate residues involved in the dimeric interfaces of RAV-1 IN CCD (novel) and RAV-1 IN CCD_{A182T} (canonical), respectively. Red stars indicate the three invariant catalytic residues, while residue A182 of RAV-1 is highlighted in cyan. Figures 1A and 1B were generated with PyMOL [60] and ESPript [61], respectively.

doi:10.1371/journal.pone.0023032.g001

antiretroviral drug that targets the catalytic site of INs (PDB entry 3L2T). It appeared that the coordinated side chain of the symmetry-related His198 occupies the position of two chelating oxygen atoms linked to the pyrimidine group of the IN inhibitor. The same observation was made in a comparison with PFV IN complexed with Mg(II) and elvitegravir, another antiretroviral molecule (PDB entry 3L2U). Again, the IN inhibitors occupy the position devoted to His198 that mimics the coordinated nucleotide in the crystal structure of RAV-1 IN CCD.

A new dimeric assembly

The canonical dimeric interface of RSV-A IN CCD (equivalent to RAV-1 IN CCD_{A182T}) which generally involves two pairs of facing α -helices of each monomer (pairs $\alpha 1/\alpha 5$; Figure 1A, right), is not visible in the crystal packing of RAV-1 IN CCD. The present A/B dimer buries three pairs of facing helices ($\alpha 3A/\alpha 5B$, $\alpha 1A/\alpha 1B$, $\alpha 3B/\alpha 5A$; Figure 1A, left) in a new intermolecular interface, which can be obtained from the crystallographic dimer of RSV-A IN CCD by a 15 Å translation of one monomer along the other. Thus, helix $\alpha 5$ faces helix $\alpha 1$ of the complementary monomer in RSV-A IN CCD, and helix $\alpha 3$ of the complementary monomer in RAV-1 IN

CCD. Moreover, helices $\alpha 1$ of monomers A and B now run almost parallel to the non-crystallographic two-fold axis and are locked together via a buried Zn(II) that coordinates the imidazole rings of His103A and His103B (Figure 1A, left). This central Zn(II) also coordinates two water molecules in a perfectly tetrahedral coordination sphere. The area of the buried surface at the new CCD/CCD interface, 740 Å² per monomer, is similar to that previously observed in RSV-A IN; that is, 750 Å² per monomer. The distance between the active sites of the two CCDs is preserved (35 Å), as is the distance between the two CCD N-termini (25 Å), while the distance between the two CCD C-termini increases significantly (from 20 Å to 35 Å). The web server PISA [34] suggests that the new dimeric assembly is stable in solution. The novel interface buries an equal number of polar and non-polar residues (Table S1) and more than 50% of contacting residues are preserved between RAV-1 IN CCD and RSV-A IN CCD (Figure 1B). For example, the ion pair between His103 (helix $\alpha 1$) and Glu187 (helix $\alpha 5$) of the complementary monomer that was highlighted in RSV-A IN CCD [35] is substituted by an equivalent intermolecular contact between Arg137 (helix $\alpha 3$) and the same Glu187 (helix $\alpha 5$) in RAV-1 IN CCD.

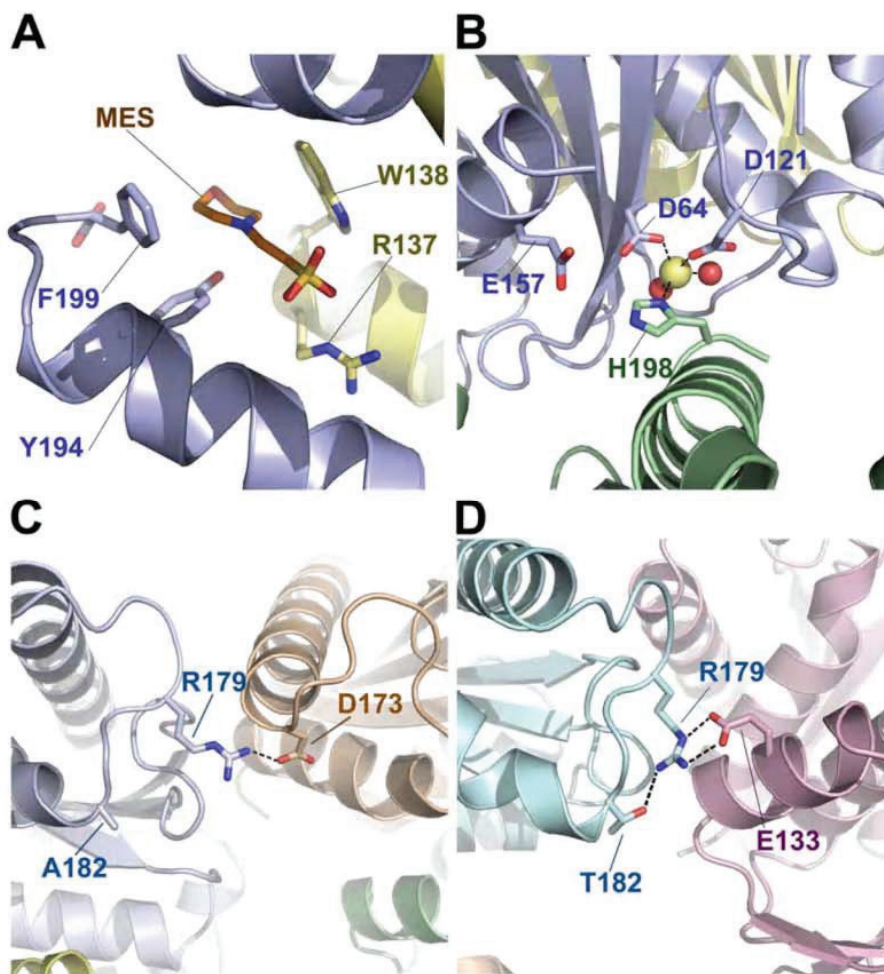


Figure 2. MES binding site and important crystallographic contacts in RAV-1 IN CCD and RAV-1 IN CCD_{A182T}. (A) The MES binding site at the novel dimeric interface of RAV-1 IN CCD. (B) A close view of the active site of RAV-1 IN CCD. The zinc ion in site I is represented by a yellow sphere. This ion is penta-coordinated by D64 and D121 (blue), H198 from a symmetry-related monomer (green) and two water molecules (red spheres). (C) A close view of RAV-1 IN CCD, focused on residue R179 near A182. Monomers are colored in light blue and pale orange (symmetry-related molecule). (D) The counterpart region for RAV-1 IN CCD_{A182T}. The A182T mutation promotes a reorientation of R179 side-chain. doi:10.1371/journal.pone.0023032.g002

Surprisingly, the substituted residue Ala/Thr 182 is not buried in any dimeric interface (canonical or novel). This residue is located on the outer edge of helix $\alpha 5$, a portion of which is accessible to solvent in both the RSV-A IN CCD and RAV-1 IN CCD crystals (Figure 1A). The A182T substitution mostly affects the side-chain orientation of the neighboring Arg179 of the $\alpha 4$ - $\alpha 5$ loop. This arginine is hydrogen bonded to the side-chain oxygen atom OG of Thr182 in RSV-A IN CCD (Figure 2D), whereas a similar contact is impossible with the aliphatic Ala182 in RAV-1 IN CCD. There, the side chain of the arginine has rotated by 130° around its CG-CD bond to mediate a crystal contact with the side-chain carboxylate in the Asp173 of a neighboring monomer (Figure 2C). Thus, the A182T substitution influences crystal assembly via Arg179 and, either large tetragonal crystals or tiny hexagonal microcrystals are observed for RAV-1 IN CCD_{A182T}, whereas only hexagonal crystals are obtained for RAV-1 IN CCD.

A buried MES molecule

The novel dimeric interface buries a MES in a canal located between helix $\alpha 5$ of molecule A and helix $\alpha 3$ of molecule B (Figure 1A, left). The O1 oxygen atom of the morpholino group is

oriented toward the bulk solvent, while the sulfonate group penetrates deeply into the interface. The morpholino group is further cradled by hydrophobic interactions with Tyr194A, Phe199A and Trp138B (Figure 2A). Its N4 nitrogen atom establishes a direct hydrogen bond with the hydroxyl group of Tyr194A, while the adjacent sulfonate group is stabilized by the guanidinium group of Arg137B. In comparison, a bound HEPES molecule is observed in the alkaline structure of RSV-A IN CCD, whereas a bound citrate is observed in the acidic structure of the same fragment [31] and in the equivalent RAV-1 IN CCD_{A182T}. However, neither of these two buffer molecules is involved in the canonical interface of RSV-A IN CCD. The HEPES molecule, which resembles MES in its sulfonate group and a six-atom cycle, is lodged at the CCD surface along the tips of loops $\beta 1\beta 2$ and $\beta 3\alpha 1$, while the citrate molecule caps the N-terminal extremity of helix $\alpha 5$ (Figure 1A, right).

The H103C mutant

In order to give proof that the novel assembly is not a crystallization artifact, a mutant able to stabilize the new dimeric interface in solution was designed. Molecular modeling shows that

the central His103 can be substituted by a cysteine to promote the formation of a disulfide bond at the new interface and covalently lock the novel quaternary structure. Thus, RAV-1 IN CCD_{H103C} was produced in specific bacteria (see ‘Materials and Methods’) and purified without β -mercaptoethanol. RAV-1 IN CCD_{H103C} was analyzed on a denaturing SDS-PAGE gel with and without reducing agent and revealed by silver staining. RAV-1 IN CCD_{H103C} migrated as a dimer in non-reducing conditions and as a monomer in reducing conditions (Figure S1). Mass spectrometry analyses were performed on reduced and non-reduced digested peptides to confirm the presence of the disulfide bond between Cys103 of monomers A and B and the formation of the novel assembly in solution.

A novel median basic groove

As described previously by Bujacz *et al.*, the canonical dimeric interface of RSV-A IN CCD contains a central cavity bordered by hydrophilic residues [35]. This cavity is conserved in HIV IN and has been investigated as a target for allosteric inhibitors [36]. A projection of molecular electrostatic potentials shows that this central invagination becomes a highly basic groove in RAV-1 IN CCD and strips the middle of the protein surface at the dimeric interface away from the catalytic sites (Figure 3A). Inspection of the narrow groove basement reveals three small pockets arranged at regular intervals so as to accommodate a linear single-stranded nucleic acid. The two outer pockets are also the gates of two symmetry-related canals running along the intermolecular interface perpendicularly to the median groove (Figure 3). The distance between the central and outer pockets is 10 Å. Residues His103A and His103B which are bridged by the central Zn(II) constitute the bottom of the middle pocket.

A ‘blind docking’ experiment was performed with a single-stranded RNA aptamer against the entire surface of our RAV-1 IN CCD dimer. All of the highest score solutions correspond to RNA fragments bound at the level of the median basic groove. According to these predictions, heterocycles of purine and pyrimidine bases could fit in the three bottom pockets of the groove with the phosphodiester chain of the RNA exposed to the solvent. The central pocket can alternatively bind to a phosphate group of the RNA backbone. In this case, the backbone adopts a linear twisted conformation, made possible by the high degree of flexibility of the chain.

Discussion

RAV-1 IN displays standard activities for 3' processing, strand transfer reactions and concerted DNA integration *in vitro* [28,37–39]. The CCD:CCD interface of IN with two pairs of facing α -helices ($\alpha 5/\alpha 1$, $\alpha 1/\alpha 5$) has been the only form observed until now. It displays a significant interaction area, the values of which range from 1500–1300 Å² for HIV-1 and SIV to 750–700 Å² for RSV and PFV. A crystal form termed II of the CCD of BIV IN was found of particular interest, because it was observed that a dimer of canonical dimers (interaction area 1250 Å²) was stabilized by a short interface named face-to-face (interaction area 580 Å²), which could also occur during the formation of the IN-DNA complex [40]. Such face-to-face interface between canonical dimers is not observed in the structure of the PFV intasome [16]. Lower to negligible interface interactions are observed for other domains (NTD or CTD) of INs [20]. Obtaining a novel dimeric arrangement for the CCD of RAV-1 IN that differs from all structures of INs published to date was a real surprise. Our first assumption was that our novel dimeric arrangement with its three pairs of facing α -helices at the interface ($\alpha 3/\alpha 5$, $\alpha 1/\alpha 1$ and $\alpha 5/\alpha 3$)

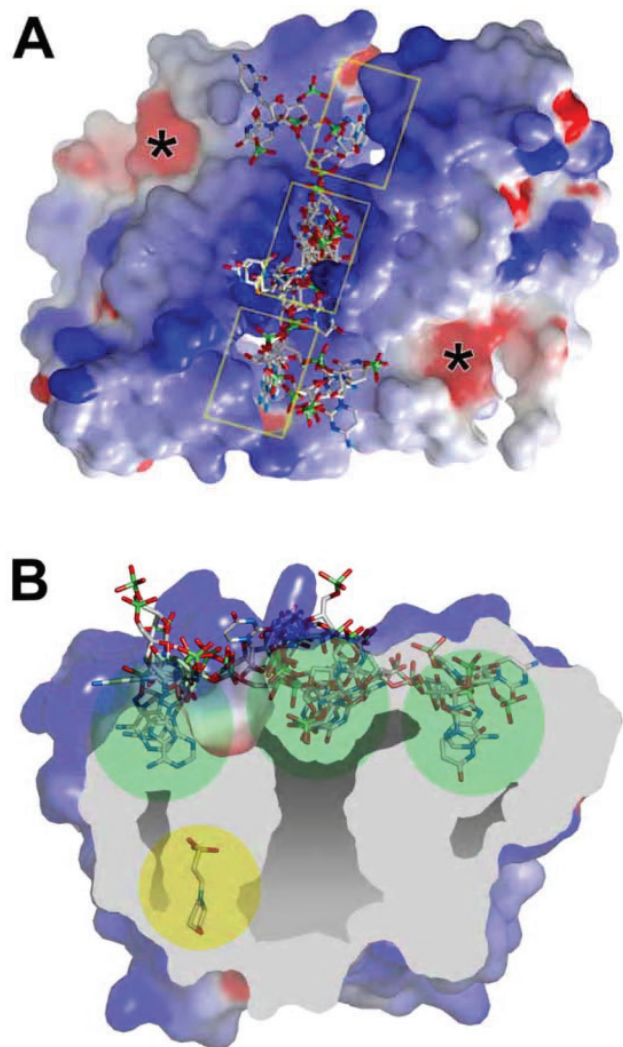


Figure 3. The novel dimer in complex with a 4-base RNA fragment, as predicted by docking experiments. (A) The molecular surface of the RAV-1 IN CCD dimer with the five top scored docking solutions (in stick representation) bound at the level of the median basic groove. Surfaces are colored according to electrostatic potential (red negative charges; blue positive charges). The three binding pockets are boxed in yellow while locations of the two active sites are indicated by black asterisk. (B) A cross-section of the molecular surface of the RAV-1 IN CCD dimer along the main axis of the basic groove demonstrating the five top scored docking conformations. Cavities and canals are colored light grey while the three binding pockets are indicated by green circles. The MES molecule buried at one canal exit is circled in yellow. The electrostatic potential map and the molecular surfaces were calculated with GRASP [62]. doi:10.1371/journal.pone.0023032.g003

was a crystallization artifact, but the formation of a covalently bonded dimer for the H103C mutant gave insight on the new dimeric assembly during bacterial production. Furthermore, the amount of buried surface in the new CCD:CCD interface is similar, ~ 750 Å², to that observed in the canonical interface of RSV-A IN, and the solvation free energy gain calculated with PISA is in agreement with the formation of a biological interaction. Zn(II) still plays an important role in the formation of the CCD-CCD interface by locking two facing histidines and may favor on its own the formation of a non-biological dimer.

However, we rather support the idea that it could be used as a cofactor to stabilize an alternate conformation of the CCD as is discussed in the following paragraphs. Avian INs are highly conserved in sequence and the dimeric association of either of the two forms can probably be achieved in most cases. As an illustration, our crystallization studies show that the two crystal forms (tetragonal and hexagonal corresponding to the canonical and novel interface, respectively) can be obtained for the RAV-1 IN CCD_{A182T} single mutant, whose sequence is identical to that of the RSV-A IN CCD peptide. Another interesting aspect of our results is to establish that the single mutation A182T, which is located away from the dimerization interface, has a considerable impact on the CCD assembly although the tertiary structure of the fragment is preserved. Most crystallographic studies of INs, especially of HIV-1 IN, were only possible after the introduction of mutations in the CCD to yield soluble proteins. In consequence, we conducted further analyses to better identify the molecular determinants of our novel interface.

The role of the pH and of the buffer

Previous studies on different types of entire INs including RAV-1 IN demonstrate that the protein exists in a monomer-dimer or a monomer-dimer-tetramer equilibrium [28,38,41]. However, our size exclusion chromatography (SEC) elution profiles of RAV-1 IN CCD and RAV-1 IN CCD_{A182T} monitored during protein purification are consistent with a monomeric protein. These results were confirmed by SEC-MALS with a concentration as high as 2.3 mg/ml (140 μ M) at sample injection (Figure S2). Consequently, we assume that the dimeric association of either of the two forms occurs during crystallization and depends on the crystallization solution.

We have tested the influence of pH on the crystal assembly of RAV-1 IN CCD, by using a wide range of pH levels from 6 to 10 in the presence of 10 mM of ZnCl₂. Crystals were obtained only at the mild acidic pH of 6 with a MES buffer. The canonical dimeric form of RSV-A IN CCD can be obtained in a citrate buffer at a similar pH [31]. Thus, an acidic solution is not the determinant of the novel quaternary structure.

The role of Zn(II)

The new interface is obtained in presence of Zn(II), which is an essential cofactor for IN because it is implicated in the folding of the NTD [12,42]. It can also be coordinated as a cofactor in two sites of the active site termed I and II [29]. Zn(II) has also been shown to stimulate the dimerization of HIV IN in association with Mg(II) [43]. Zinc binding has been studied intensively for the canonical RSV-A IN CCD [14,29]. Crystal-soaking experiments were performed in solutions containing 2 mM to 100 mM ZnCl₂, and four coordinated Zn(II) were subsequently observed at sites I and II of the catalytic pocket and at two distant sites termed III and IV. Interestingly, binding site similarities are observed between this soaked structure and the novel crystal structure of RAV-1 IN CCD. Site I is conserved in both forms, with a Zn(II) bridging the catalytic residues Asp64 and Asp121. This further demonstrates that the novel dimeric assembly has no influence on the topology of the active site. Accordingly, site II is not occupied as in most retroviral INs. Structural similarities between sites III and IV raise more questions, because the local environments of the two dimeric forms are distinct. In canonical RSV-A IN CCD, the two remaining Zn(II) are found at the surface of the protein and are coordinated by His103 (site III) and His198 (site IV), respectively. In the novel RAV-1 IN CCD structure, His103 is buried deep in the dimeric interface and is connected to His103 of the complementary monomer via one Zn(II) ion (Figure 1A).

His198 is still located at the protein surface, but it is coordinated to site I of a symmetry-related monomer via a second Zn(II) (Figure 2B). Thus, sites I and IV of RSV-A IN CCD merge into a single penta-coordinated site I in RAV-1 IN CCD, while sites III of each RSV-A IN monomer fuse into a central Zn(II) site to lock the RAV IN dimer together.

From these data, it appears that Zn(II) plays an important role in the formation of the novel interface as it does for the folding of the NTD. In our case, this divalent cation binds to the accessible His103 and His198 of the monomeric fragment and promotes dimerization and crystal growth. As an illustration, we produced and tried to crystallize the H103A mutant and the H103A/A182T double mutant of RAV-1 IN CCD using the “new” and the “canonical” crystallization conditions. No crystal was obtained with RAV-1 IN CCD_{H103A}, and large crystals of RAV-1 IN CCD_{H103A/A182T} were observed but with the “canonical” condition only. These results, like those obtained with the H103C mutant, suggest that both His103 and a bound Zn(II) are necessary to the creation of the new interface.

We also tried to characterize the emergence of the novel quaternary assembly in solution by Dynamic Light Scattering (DLS) and Small Angle X-ray Scattering (SAXS). However, the CCD domain of RAV-1 IN tends to aggregate in the presence of zinc so no reliable measures were obtained.

The putative biological role of the novel interface

The biological relevance of the novel dimeric form should now be questioned with respect to the retroviral cycle. Although the C α trace of each CCD monomer is preserved, the novel association might result in significant displacements of the two terminal domains fused to the CCD, affecting the entire protein. However, *in vitro* strand transfer reactions within the entire RAV-1 IN protein are optimal at alkaline pH in the absence of zinc [37]. These experimental conditions are the opposite of those required for the formation of the novel interface *in crystallo*, and we believe that a biological function should be investigated apart from the integration mechanism. IN exhibits karyophilic properties and one can also propose that the basic groove located at the novel interface could be used as a karyophilic determinant. This motif could complement the nuclear localization signal (NLS) that was identified in the region linking the CCD and the CTD of RSV IN [44,45].

IN is also involved in reverse transcription and virus assembly, as shown for HIV-1 [46–48]. These two steps of the viral cycle occur in presence of viral RNA, and the potential ability of the new basic groove to bind a single-stranded nucleic acid chain has enabled us to suggest possible biological roles for the novel interface. ASLV reverse transcriptase (RT) is an $\alpha\beta$ heterodimer, which contains the polymerase, RNase H, and IN domains within the 95 kDa β subunit. Cleavage of the IN domain from the β subunit produces the 63 kDa α subunit and free IN enzyme. One function of the IN domain in the β subunit is to increase the affinity of RT to its substrate [49,50]. Therefore, we speculate that this new quaternary structure may contribute to the binding of viral RNA or the single-stranded strong stop DNA generated during reverse transcription.

Finally, this dimeric structure shows that the binding of a MES morpholino group within the new interface has induced an important movement of residue F199 at the C-terminus of helix α 5. This residue is spatially equivalent to F185 in HIV IN, which is often mutated to lysine or histidine in crystallographic studies in order to increase protein solubility [23,51]. Such a substitution may impair MES fixation and could have hindered the detection of the novel dimeric interface in HIV IN.

Anyhow, our findings suggest that avian IN CCDs may have at least two intermolecular interfaces permitting multifunctionality. A parallel could be drawn with other retroviral proteins, such as Vif in lentivirus, which contain intrinsically disordered regions and can therefore interact with multiple partners [52]. Retroviruses with limited genome length could use this strategy to generate proteins with flexible structures to mediate more than one step of the viral cycle.

Materials and Methods

Cloning the RAV-1 IN CCD sequence: pETG10a-INRAV1_{CCD}

The DNA sequence encoding the IN catalytic core domain of RAV-1 (residues 53–199) was amplified by PCR from the pET30a-INRAV1 plasmid [37]. The sequence has been deposited in GenBank nucleotide database under accession number JF514545. The fragment was cloned using Gateway Technology (Invitrogen); the 5' *attB* PCR primer was designed with a thrombin cleavage site. pDONR223 was used as donor vector to generate an entry clone. pETG10a, containing a hexahistidine tag, was used as the expression vector. The constructed expression vector was confirmed by DNA sequencing.

Site-directed mutagenesis

The pETG10a-INRAV1_{CCD}A182T, pETG10a-INRAV1_{CCD}-H103A, pETG10a-INRAV1_{CCD}H103A/A182T, pETG10a-INRAV1_{CCD}H103C and pET30a-INRAV1-H130C sequences were created by site-directed mutagenesis (Stratagene QuickChange kit) using pETG10a-INRAV1_{CCD} and pET30a-INRAV1 as a template for the following mutagenic primers (mutations are shown as lowercase letters): 5'-GTTCCCCCTGTTTGGCTggtGGGGATTCTTTTCATAAAG-3' (F-A182T), 5'-CTTTATGAAAAGAA-TCCCCaccAGCAAACAGGGGGAAC-3' (R-A182T); 5'-GC-CGTGGCCCAATGagcTTGTGCAGCAACCG-3' (F-H103A), 5'-CGGTTGCTGCACAAgctCATTTGGGCCACGGC-3' (R-H103A) and 5'-GCCGTGGCCCAATGacaTTGTGCAGCAACCG-3' (F-H103C), 5'-CGGTTGCTGCACAAgctCATTTGGGCCACGGC-3' (R-H103C). Mutations to the expression vector were confirmed by DNA sequencing. pETG10a-INRAV1_{CCD}, pETG10a-INRAV1_{CCD}A182T, pETG10a-INRAV1_{CCD}H103A and pETG10a-INRAV1_{CCD}H103A/A182T were introduced into *E. coli* BL21 (DE3) pLysS competent cells (Novagen) for protein expression. pETG10a-INRAV1_{CCD}H103C was introduced in *E. coli* Rosetta-gamiTM B(DE3)pLysS competent cells (Novagen) for protein expression and formation of target protein disulphide bond in the bacterial cytoplasm.

Expression and purification of RAV-1 IN CCD and RAV-1 IN CCD mutants (A182T, H103A, H103A/A182T)

An overnight culture (5 ml) from a single colony containing the desired plasmid was used to inoculate 1 l fresh LB medium in the presence of ampicillin (50 µg/ml). The culture was incubated at 37°C, and shaken at 220 rpm until an A_{600} = 0.8–0.9 was reached. Overexpression of the proteins was induced by 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) overnight at 25°C. Then, bacteria were harvested by centrifugation at 3000 × g for 10 min and stored at –80°C. For purification, the thawed bacterial pellet was sonicated in 20 ml of buffer A (0.5 M NaCl, 10 mM imidazole, 5 mM β-mercaptoethanol, 20 mM Tris-HCl, pH 8.0) in the presence of 100 µl Halt Protease Inhibitor Cocktail (Pierce) and DNase/RNase (final concentration 50 µg/ml). The lysate was cleared by centrifugation (45 min, 10,000 rpm at 4°C) and then filtered through a 0.45 µm filter. The supernatant

containing soluble His-tagged proteins was loaded on an Ni²⁺ charged 1 ml HiTrap Chelating HP column (GE Healthcare) using the ÄKTA chromatography system. The column was extensively washed with buffer B (0.5 M NaCl, 20 mM imidazole, 5 mM β-mercaptoethanol, 20 mM Tris-HCl, pH 8.0) and with 5 ml Buffer B2 (1 M NaCl, 20 mM imidazole, 5 mM β-mercaptoethanol, 20 mM Tris-HCl, pH 8.0). IN proteins were eluted with 0.5 M NaCl, 500 mM imidazole, 5 mM β-mercaptoethanol, 20 mM Tris-HCl, pH 8.0 using a linear gradient. Eluted fractions were collected and analysed by SDS-PAGE. The hexahistidine tag was removed by overnight digestion with thrombin protease (Amersham Biosciences) at 4°C. Digested protein solutions were loaded on a Ni²⁺ charged 1 ml HiTrap Chelating HP column (GE Healthcare); undigested proteins and free tags were fixed on the column while digested protein was recovered in the flow through. The purity of the recovered protein was analyzed by SDS-PAGE and silver staining. Protein concentration was determined according to a Bradford assay (Bio-Rad Laboratories) using BSA as standard and then concentrated to 10 mg/ml using a 10 kDa molecular-weight cut-off membrane (Vivascience).

Crystallization of RAV-1 IN CCD and “canonical” RAV-1 IN CCD_{A182T}

Crystallization conditions were searched for RAV-1 IN CCD using the sitting-drop vapour-diffusion method and commercial kits from Hampton Research, Molecular Dimensions Limited (MDL) and Qiagen. Droplets composed of 0.3 µl protein solution at 10 mg/ml and an equal volume of crystallization solution were equilibrated against 100 µl reservoir solution within a sealed well at 18°C. A crystal was observed for condition 24 of MDL PACT *premier* (10 mM ZnCl₂, 20% (w/v) PEG 6000, 100 mM MES, pH 6.0). Crystals reached maximum dimensions of 20 × 20 × 5 µm³ within a week. Their size was improved to 150 × 150 × 80 µm³ with a macroseeding technique using drops containing a 2 µl protein solution and a 2 µl precipitant solution. Crystals were mounted in a nylon loop and cryoprotected by adding 0.4 µl ethylene glycol to the hanging drop before flash-freezing in liquid nitrogen. Crystals of RAV-1 IN CCD_{A182T} were obtained in hanging drops at 18°C by mixing 1 µl protein solution (5 mg/ml) and 1 µl reservoir solution with either the crystallization condition of the wild-type or a crystallization condition for RSV-A IN CCD (20% (w/v) PEG 4000, 10% isopropanol and 0.1 M Na citrate pH 6.2). Micro-crystals were obtained under the first set of conditions. Large crystals with maximum dimensions of 200 × 150 × 150 µm³ were obtained under the second set of conditions.

Data collection and structure determination

A synchrotron data set for RAV-1 IN CCD was collected to 1.8 Å resolution from a crystal cooled to 100 K at ESRF beamline ID29 (Grenoble, France) at a wavelength of 1.28 Å. Data were processed with XDS/XSCALE [53]. The phase problem was solved by molecular replacement using the program AMoRe [54]. The final crystal structure, with two polypeptide chains named A and B, was obtained by alternating cycles of restrained refinement in Refmac5 [55] and manual rebuilding in Coot [56]. WHATCHECK [57] was used to assess the geometric quality of the model (94.2% of the residues in the most favored region of the Ramachandran plot). A second synchrotron data set was collected to 1.55 Å resolution at the ESRF beamline BM30A at a wavelength of 0.98 Å, under cryo-conditions (100 K), from a crystal of RAV-1 IN CCD_{A182T}. The phase problem was solved by rigid-body refinement with Refmac5 prior to restrained refinement. Additional data collection and refinement statistics are

Table 1. Data collection and refinement statistics.

	RAV-1 IN CCD	"Canonical" RAV-1 IN CCD _{A182T}
Data collection		
Space group	P6 ₁	P4 ₃ 2 ₁ 2
Cell dimensions:		
a, b, c (Å)	107.5, 107.5, 50.9	65.5, 65.5, 79.2
α, β, γ (°)	90, 90, 120	90, 90, 90
V_m (Å ³ .Da ⁻¹)	2.6	2.8
Solvent content (%)	52	55
Resolution range (Å)	19.7–1.8 (1.95–1.8)	20.0–1.55 (1.60–1.55)
Total number of reflections	162065 (33802)	111679 (7650)
Number of unique reflections	30860 (6613)	24970 (2171)
R_{sym} (%)	7.6 (39.9)	4.9 (25.6)
$I/\sigma(I)$	11.6 (3.4)	20.9 (5.4)
Completeness (%)	99.0 (99.0)	97.4 (95.3)
Redundancy	5.3 (5.1)	4.5 (3.5)
Refinement		
Resolution (Å)	1.8	1.55
R_{factor}/R_{free} (%)	19.3/22.8	16.4/21.1
No. atoms:		
protein (non-hydrogen)	2099	1079
ligand/ion	15	13
Water	186	122
Overall B-factor (Å ²)	31.7	15.2
R.m.s. deviations:		
bond lengths (Å)	0.020	0.026
bond angles (°)	2.34	1.72

Values in parentheses are for highest-resolution shell.
doi:10.1371/journal.pone.0023032.t001

presented in Table 1 for both structures. Coordinates and structure factors of crystal structures described herein have been deposited in the RCSB Protein DataBank (<http://www.rcsb.org>) under the accession codes 3O4N (RAV-1 IN CCD) and 3O4Q (RAV-1 IN CCD_{A182T} mutant).

Docking experiments

A single-stranded RNA molecule comprising four bases was designed. The number of bases employed was determined by a computational limit to the number of rotatable bonds allowed. The 1.1 Å atomic resolution crystal structure of the DNA octanucleotide d(pATTCATTC) was used as the template (PDB entry 284D). We truncated and modified this structure to obtain our final RNA fragment, pCAUUp. This ligand and the receptor structure of RAV-1 IN CCD were then prepared with AutoDockTools [58]. A "blind docking" was subsequently carried out on the entire surface of the dimer with the program AutoDock Vina [59]. Once a binding area was identified, new docking cycles were achieved with a reduced search space encompassing the site of interest. Flexible-ligand docking with grid-based energy scoring was conducted with the program's standard protocol [59].

Supporting Information

Figure S1 Visualization of disulfide bonds in RAV-1 IN CCD_{H103C} by SDS-PAGE in non-reducing and reducing conditions. RAV-1 IN CCD_{H103C} was produced in *E. coli*

Rosetta-gamiTM B(DE3)pLysS competent cells (Novagen) as described in 'Materials and Methods'. Track 1: RAV-1 IN in reducing conditions (β -mercaptoethanol 5%). A single band corresponding to the monomeric form is observed (theoretical molecular weight of 16.3 kDa). Track 2: RAV-1 IN CCD_{H103C} in non-reducing conditions. Monomeric and dimeric forms are observed (theoretical MW of 16.3 kDa and 32.6 kDa, respectively). The apparition of the high molecular weight strip attests the production of dimeric RAV-1 IN CCD_{H103C} with disulfide bonds. Track 3: molecular weight markers. RAV-1 IN CCD_{H103C} was loaded onto a 12% SDS PAGE in reducing and non-reducing conditions, and the protein bands were detected by Coomassie Blue Staining. The two bands corresponding to RAV-1 IN CCD_{H103C} with or without a putative intramolecular disulfide bond were excised and cut to perform in gel trypsin digestion without reduction and alkylation [63]. The tryptic peptides were analysed by MALDI-TOF and mass spectra were recorded on a Voyager DE-PRO (AB Sciex). (TIF)

Figure S2 SEC-MALS/RI analysis of RAV-1 IN CCD in solution. Determination of the oligomerization state of RAV-1 IN CCD in solution was studied by the combination of UV spectrometry, multi-angle static light scattering (MALS), and refractometry, coupled on-line with an analytical size exclusion chromatography (SEC) column. UV, MALS and refractometry measurements were achieved with a Photo Diode Array 2996

(Waters), a MiniDawn Treos (Wyatt Technology), and an Optilab rEX (Wyatt Technology), respectively. Size exclusion chromatography was carried out on an Alliance 2695 HPLC system (Waters) using a KW803 column (Shodex) run in a buffer containing 20 mM Tris-HCl, 500 mM NaCl and 5 mM β -Mercaptoethanol at pH 7.5 with a flow rate of 0.5 ml/min. The molar mass (left axis, bold line) and the UV_{280 nm} absorbance (right axis, regular line) are plotted as a function of the column elution volume. SEC-MALS/RI/UV characterization revealed a mass of 19800 ± 210 g/mol and 17690 ± 190 g/mol, respectively. These data attest the monomeric nature of the protein.

(TIF)

Table S1 Residues involved in the novel dimeric interface. (DOC)

References

- Lewinski MK, Bushman FD (2005) Retroviral DNA integration—mechanism and consequences. *Adv Genet* 55: 147–181.
- Pommier Y, Neamati N (1999) Inhibitors of human immunodeficiency virus integrase. *Adv Virus Res* 52: 427–458.
- Hazuda DJ, Young SD, Guare JP, Anthony NJ, Gomez RP, et al. (2004) Integrase inhibitors and cellular immunity suppress retroviral replication in rhesus macaques. *Science* 305: 528–532.
- Sato M, Motomura T, Aramaki H, Matsuda T, Yamashita M, et al. (2006) Novel HIV-1 integrase inhibitors derived from quinolone antibiotics. *J Med Chem* 49: 1506–1508.
- Marchand C, Maddali K, Metifiot M, Pommier Y (2009) HIV-1 IN inhibitors: 2010 update and perspectives. *Curr Top Med Chem* 9: 1016–1037.
- Skalka AM, Katz RA (2005) Retroviral DNA integration and the DNA damage response. *Cell Death Differ* 12 Suppl 1: 971–978.
- Smith JA, Daniel R (2006) Following the path of the virus: the exploitation of host DNA repair mechanisms by retroviruses. *ACS Chem Biol* 1: 217–226.
- Smith JA, Wang FX, Zhang H, Wu KJ, Williams KJ, et al. (2008) Evidence that the Nijmegen breakage syndrome protein, an early sensor of double-strand DNA breaks (DSB), is involved in HIV-1 post-integration repair by recruiting the ataxia telangiectasia-mutated kinase in a process similar to, but distinct from, cellular DSB repair. *Virology* 375: 11.
- Zhao Z, McKee CJ, Kessl JJ, Santos WL, Daigle JE, et al. (2008) Subunit-specific protein footprinting reveals significant structural rearrangements and a role for N-terminal Lys-14 of HIV-1 Integrase during viral DNA binding. *J Biol Chem* 283: 5632–5641.
- Heuer TS, Brown PO (1997) Mapping features of HIV-1 integrase near selected sites on viral and target DNA molecules in an active enzyme-DNA complex by photo-cross-linking. *Biochemistry* 36: 10655–10665.
- Heuer TS, Brown PO (1998) Photo-cross-linking studies suggest a model for the architecture of an active human immunodeficiency virus type 1 integrase-DNA complex. *Biochemistry* 37: 6667–6678.
- Zheng R, Jenkins TM, Craigie R (1996) Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity. *Proc Natl Acad Sci USA* 93: 13659–13664.
- Bujacz G, Jaskolski M, Alexandratos J, Wlodawer A, Merkel G, et al. (1996) The catalytic domain of avian sarcoma virus integrase: conformation of the active-site residues in the presence of divalent cations. *Structure* 4: 89–96.
- Bujacz G, Alexandratos J, Wlodawer A, Merkel G, Andrade M, et al. (1997) Binding of different divalent cations to the active site of avian sarcoma virus integrase and their effects on enzymatic activity. *J Biol Chem* 272: 18161–18168.
- Leh H, Brodin P, Bischerour J, Deprez E, Tauc P, et al. (2000) Determinants of Mg²⁺-dependent activities of recombinant human immunodeficiency virus type 1 integrase. *Biochemistry* 39: 9285–9294.
- Hare S, Gupta SS, Valkov E, Engelman A, Cherepanov P (2010) Retroviral integrase assembly and inhibition of DNA strand transfer. *Nature* 464: 232–236.
- Bushman FD, Engelman A, Palmer I, Wingfield P, Craigie R (1993) Domains of the integrase protein of human immunodeficiency virus type 1 responsible for polynucleotidyl transfer and zinc binding. *Proc Natl Acad Sci USA* 90: 3428–3432.
- Murzin AG, Brenner SE, Hubbard T, Chothia C (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* 247: 536–540.
- Nowotny M (2009) Retroviral integrase superfamily: the structural perspective. *EMBO Rep* 10: 144–151.
- Jaskolski M, Alexandratos JN, Bujacz G, Wlodawer A (2009) Piecing together the structure of retroviral integrase, an important target in AIDS therapy. *FEBS J* 276: 2926–2946.
- Engelman A, Hickman AB, Craigie R (1994) The core and carboxyl-terminal domains of the integrase protein of human immunodeficiency virus type 1 each contribute to nonspecific DNA binding. *J Virol* 68: 5911–5917.
- Andrake MD, Skalka AM (1995) Multimerization determinants reside in both the catalytic core and C terminus of avian sarcoma virus integrase. *J Biol Chem* 270: 29299–29306.
- Chen JC, Krucinski J, Miercke IJ, Finer-Moore JS, Tang AH, et al. (2000) Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding. *Proc Natl Acad Sci USA* 97: 8233–8238.
- Wang JY, Ling H, Yang W, Craigie R (2001) Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein. *EMBO J* 20: 7333–7343.
- Yang ZN, Mueser TC, Bushman FD, Hyde CC (2000) Crystal structure of an active two-domain derivative of Rous sarcoma virus integrase. *J Mol Biol* 296: 535–548.
- Michel F, Crucifix C, Granger F, Eiler S, Mouscadet JF, et al. (2009) Structural basis for HIV-1 DNA integration in the human genome, role of the LEDGF/p75 cofactor. *EMBO J* 28: 980–991.
- Maertens GN, Hare S, Cherepanov P (2010) The mechanism of retroviral integration from X-ray structures of its key intermediates. *Nature* 468: 326–329.
- Moreau K, Faure C, Violot S, Gouet P, Verdier G, et al. (2004) Mutational analyses of the core domain of Avian Leukemia and Sarcoma Viruses integrase: critical residues for concerted integration and multimerization. *Virology* 318: 566–581.
- Lubkowski J, Yang F, Alexandratos J, Merkel G, Katz RA, et al. (1998) Structural basis for inactivating mutations and pH-dependent activity of avian sarcoma virus integrase. *J Biol Chem* 273: 32685–32689.
- Lubkowski J, Yang F, Alexandratos J, Wlodawer A, Zhao H, et al. (1998) Structure of the catalytic domain of avian sarcoma virus integrase with a bound HIV-1 integrase-targeted inhibitor. *Proc Natl Acad Sci USA* 95: 4831–4836.
- Lubkowski J, Dauter Z, Yang F, Alexandratos J, Merkel G, et al. (1999) Atomic resolution structures of the core domain of avian sarcoma virus integrase and its D64N mutant. *Biochemistry* 38: 13512–13522.
- Greenwald J, Le V, Butler SL, Bushman FD, Choe S (1999) The mobility of an HIV-1 integrase active site loop is correlated with catalytic activity. *Biochemistry* 38: 8892–8898.
- Dias A, Bouvier D, Crepin T, McCarthy AA, Hart DJ, et al. (2009) The capsid-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature* 458: 914–918.
- Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. *J Mol Biol* 372: 774–797.
- Bujacz G, Jaskolski M, Alexandratos J, Wlodawer A, Merkel G, et al. (1995) High-resolution structure of the catalytic domain of avian sarcoma virus integrase. *J Mol Biol* 253: 333–346.
- Maroun RG, Gayet S, Benleulmi MS, Porumb H, Zargarian L, et al. (2001) Peptide inhibitors of HIV-1 integrase dissociate the enzyme oligomers. *Biochemistry* 40: 13840–13848.
- Moreau K, Faure C, Verdier G, Ronfort C (2002) Analysis of conserved and non-conserved amino acids critical for ALSV (Avian leukemia and sarcoma viruses) integrase functions *in vitro*. *Arch Virol* 147: 1761–1778.
- Moreau K, Faure C, Violot S, Verdier G, Ronfort C (2003) Mutations in the C-terminal domain of ALSV (Avian Leukemia and Sarcoma Viruses) integrase alter the concerted DNA integration process *in vitro*. *Eur J Biochem* 270: 4426–4438.
- Moreau K, Charmentant J, Gallay K, Faure C, Verdier G, et al. (2009) Avian sarcoma and leukemia virus (ASLV) integration *in vitro*: mutation or deletion of integrase (IN) recognition sequences does not prevent but only reduces the efficiency and accuracy of DNA integration. *Virology* 392: 94–102.
- Yao X, Fang S, Qiao W, Geng Y, Shen Y (2010) Crystal structures of catalytic core domain of HIV-1 integrase: implications for the interaction between integrase and target DNA. *Protein Cell* 1: 363–370.
- Coleman J, Eaton S, Merkel G, Skalka AM, Laue T (1999) Characterization of the self association of Avian sarcoma virus integrase by analytical ultracentrifugation. *J Biol Chem* 274: 32842–32846.

Acknowledgments

We are grateful to the staff members of ID29 and FIP-BM30A beamlines at the ESRF synchrotron (Grenoble, France) for technical advices and precious help during data collection. We also thank Dr M. Becchi and the CCMP (Centre Commun de Microanalyse des Protéines) for the mass spectrometry analyses, Dr A. Chaboud and PAP platform (Production et Analyse de Protéines) for the technical support. We thank N. Boudehen for her participation in cloning.

Author Contributions

Conceived and designed the experiments: AB PG. Performed the experiments: AB KM XR M-PC RM RH CR PG. Analyzed the data: AB KM XR M-PC RM CR PG. Contributed reagents/materials/analysis tools: AB KM XR M-PC RM CR PG. Wrote the paper: AB XR CR PG.

42. Cai M, Zheng R, Caffrey M, Craigie R, Clore GM, et al. (1997) Solution structure of the N-terminal zinc binding domain of HIV-1 integrase. *Nat Struct Biol* 4: 567–577.
43. Deprez E, Tauc P, Leh H, Mouscadet JF, Auclair C, et al. (2000) Oligomeric states of the HIV-1 integrase as measured by time-resolved fluorescence anisotropy. *Biochemistry* 39: 9275–9284.
44. Kukulj G, Jones KS, Skalka AM (1997) Subcellular localization of avian sarcoma virus and human immunodeficiency virus type 1 integrases. *J Virol* 71: 843–847.
45. Andrade MD, Sauter MM, Boland K, Goldstein AD, Hussein M, et al. (2008) Nuclear import of Avian Sarcoma Virus integrase is facilitated by host cell factors. *Retrovirology* 5: 73.
46. Engelman A, Englund G, Orenstein JM, Martin MA, Craigie R (1995) Multiple effects of mutations in human immunodeficiency virus type 1 integrase on viral replication. *J Virol* 69: 2729–2736.
47. Zhu K, Dobard C, Chow SA (2004) Requirement for integrase during reverse transcription of human immunodeficiency virus type 1 and the effect of cysteine mutations of integrase on its interactions with reverse transcriptase. *J Virol* 78: 5045–5055.
48. Nishitsuji H, Hayashi T, Takahashi T, Miyano M, Kannagi M, et al. (2009) Augmentation of reverse transcription by integrase through an interaction with host factor, SIP1/Gemin2 Is critical for HIV-1 infection. *PLoS One* 4: e7825.
49. Soltis DA, Skalka AM (1988) The alpha and beta chains of avian retrovirus reverse transcriptase independently expressed in *Escherichia coli*: characterization of enzymatic activities. *Proc Natl Acad Sci USA* 85: 3372–3376.
50. Werner S, Wohl BM (1999) Soluble Rous sarcoma virus reverse transcriptases alpha, alphabeta, and beta purified from insect cells are processive DNA polymerases that lack an RNase H 3'→5' directed processing activity. *J Biol Chem* 274: 26329–26336.
51. Jenkins TM, Engelman A, Ghirlando R, Craigie R (1996) A soluble active mutant of HIV-1 integrase: involvement of both the core and carboxyl-terminal domains in multimerization. *J Biol Chem* 271: 7712–7718.
52. Reingewertz TH, Shalev DE, Friedler A (2010) Structural Disorder in the HIV-1 Vif Protein and Interaction-Dependent Gain of Structure. *Protein Pept Lett* 17: 988–998.
53. Kabsch W (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Cryst* 26: 795–800.
54. Navaza J (2001) Implementation of molecular replacement in AMoRe. *Acta Crystallogr D* 57: 1367–1372.
55. CCP4 (1994) Collaborative Computational Project Number 4. *Acta Crystallogr D* 50: 760–763.
56. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D* 60: 2126–2132.
57. Hoof RW, Vriend G, Sander C, Abola EE (1996) Errors in protein structures. *Nature* 381: 272.
58. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, et al. (1998) Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *J Computational Chemistry* 19: 1639–1662.
59. Trott O, Olson AJ (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 31: 455–461.
60. DeLano WL (2008) The PyMOL Molecular Graphics System. Palo Alto, CA, USA: DeLano Scientific LLC.
61. Gouet P, Robert X, Courcelle E (2003) ESPript/ENDscript: Extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucl Acids Res* 31: 3320–3323.
62. Nicholls A, Sharp KA, Honig B (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* 11: 281–296.
63. Shevchenko A, Wilm M, Vorm O, Mann M (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem* 68: 850–858.

PARTIE 2 :
ETUDE CRISTALLOGRAPHIQUE D'ALPHA-
GALACTOSIDASES THERMOSTABLES ISSUES DU
MICROORGANISME *GEOBACILLUS*
STEAROTHERMOPHILUS

EXPOSÉ BIBLIOGRAPHIQUE

1. Généralités

Les glucides sont des molécules organiques contenant au minimum un groupement carbonyle (aldéhyde ou cétone) et plusieurs hydroxyles. Ces molécules peuvent être complexes de par les nombreuses combinaisons possibles de polymérisation des unités de sucres simples appelées « oses », et de la nature des liaisons, permettant l'assemblage de ces sucres simples en sucres complexes ou « osides ». De plus, les glucides vont également pouvoir incorporer différents hétéroatomes comme le phosphore, l'azote ou encore le soufre augmentant cette diversité.

Les glucides ont d'importants rôles biologiques qui peuvent être classés en trois grandes catégories : le stockage d'énergie, un rôle structural et la communication cellulaire et les interactions cellule-cellule (Stick & Williams 2008).

Le rôle de stockage d'énergie est d'une importance majeure chez les plantes ainsi que chez les animaux. En particulier, ce stockage pourra être réalisé sous la forme de polymères ramifiés d'unités de glucose comme l'amidon chez les plantes ou le glycogène chez les animaux. L'hydrolyse de ces polymères catalysée par des enzymes spécifiques nommées « glycoside hydrolases », assurera la libération d'unités de glucose, qui permettra de répondre rapidement aux besoins énergétiques des cellules.

Les glucides possèdent également un rôle structural important. La cellulose, qui est un polymère linéaire de glucose, constitue une composante majeure de la paroi cellulaire végétale. Un autre exemple est celui de la chitine, qui est un polymère linéaire d'unité de N-acétyl-glucosamine et qui participe à la formation de l'exosquelette chez les crustacés, les insectes ainsi que chez certains autres organismes. Chez les animaux, les protéoglycanes constitués de la combinaison d'une protéine avec des polymères linéaires de sucres (les glycosaminoglycanes) constituent une composante essentielle de la matrice extracellulaire qui joue un rôle de soutien structural pour les tissus.

Enfin, ils vont jouer un rôle dans les interactions cellule-cellule et dans la communication cellulaire. Les sucres impliqués dans ces fonctions sont principalement retrouvés sous la forme de glycoconjugués. Les glycoconjugués sont constitués d'oligosaccharides associés covalamment à des protéines (glycoprotéines) ou à des lipides (glycolipides). Ils sont majoritairement retrouvés à la surface des cellules ou dans les espaces extra-cellulaires et vont pouvoir interagir avec certaines protéines fixant les sucres (cytokines, lectines...) participant ainsi aux processus de communication cellulaire. Ces molécules, qui participent ainsi à des processus physiologiques, sont également impliquées dans certains processus pathologiques et peuvent contribuer à l'adressage et à l'adhésion de pathogènes bactériens et viraux à la surface des cellules.

Les glycoside hydrolases sont des enzymes clés du métabolisme des sucres et se trouvent chez les archéobactéries, les eubactéries, les eucaryotes mais également les virus. Elles constituent un très large groupe d'enzymes capable de catalyser l'hydrolyse de liaisons glycosidiques entre deux glucides ou un glucide et un groupement non glucidique ou groupement aglycone. Certaines de ces enzymes seront également capables sous certaines conditions expérimentales de catalyser des réactions de synthèse de liaisons glycosidiques (Perugino *et al.* 2004).

2. Classification

La grande diversité des glycoside hydrolases a nécessité la mise en place de systèmes de classification basés sur différents critères.

Les glycoside hydrolases sont des enzymes, et leurs activités sont classifiées par la nomenclature de l'Union Internationale de Biochimie et de Biologie Moléculaire (IUBMB) depuis 1984. Cette classification est basée sur le type de réaction catalysée par l'enzyme mais également sur sa spécificité de substrat, et permet de donner à l'enzyme un numéro EC ou *Enzyme Commission number*. La nomenclature EC utilise quatre nombres, le premier indique le type de réaction catalysée, alors que les deuxième, troisième et quatrième nombres précisent la nature du substrat impliqué dans la réaction (Webb 1992). Les glycoside hydrolases appartiennent dans la nomenclature EC au groupe EC 3.2. Ce groupe regroupe l'ensemble des enzymes à activité hydrolase (EC 3) et qui vont agir spécifiquement sur les liaisons glycosidiques (EC 3.2). Des glycoside hydrolases pourront présenter une spécificité pour les liaisons O et S-glycosidiques (EC 3.2.1), N-glycosidiques (EC 3.2.2) ou uniquement S-glycosidiques (EC 3.2.3).

Ce classement a vite montré ses limites avec l'avancée des études biochimiques et structurales sur ces enzymes. En effet, cette classification ne tient pas compte des similarités de séquences protéiques entre ces enzymes ainsi que de leurs caractéristiques structurales. Ce classement ne tient également pas compte de la nature de la réaction catalysée par l'enzyme (réactions d'inversion ou de rétention chez les glycoside hydrolases, participation de cofacteurs) et du fait que ces enzymes peuvent agir sur différents substrats.

Aussi, une nouvelle classification a été établie et constitue la classification de B. Henrissat disponible sur la banque de données CAZy (pour *Carbohydrate Active enZymes*). Cette classification, proposée en 1991, va regrouper les glycoside hydrolases en familles en fonction des homologies de leur séquence d'acides aminés (Henrissat 1991; Henrissat & Bairoch 1993). Chaque famille possèdera généralement un repliement et un mécanisme catalytique propre et ce, même si ses membres agissent sur différents substrats ou proviennent d'organismes très éloignés en terme d'évolution. Comme le repliement tridimensionnel des protéines est généralement mieux conservé que la séquence, certaines de ces familles sont regroupées en clans ou superfamilles en fonction de la conservation de leur structure tridimensionnelle et de leur similarité mécanistique (Henrissat & Bairoch 1996).

Cette classification est régulièrement réactualisée, et il y a actuellement 132 familles (dont cinq ne font toutefois plus partie de cette classification) de glycoside hydrolases ainsi que quatorze clans (GH-A à GH-N). Cette classification permet donc de prendre en considération la structure tridimensionnelle ainsi que le mécanisme moléculaire des glycoside hydrolases (Cantarel *et al.* 2009).

3. Mécanismes catalytiques

Les glycoside hydrolases vont majoritairement catalyser l'hydrolyse de liaisons glycosidiques. Ces enzymes sont toutefois également capables de catalyser la formation de liaisons glycosidiques en catalysant des réactions d'hydrolyse inverse ou encore de transglycosylation. La réaction d'hydrolyse constitue la réaction majoritaire en milieu aqueux, cependant, certaines conditions expérimentales vont favoriser les réactions d'hydrolyse inverse et de transglycosylation, en particulier par la présence d'un excès de substrat ou encore par la réduction de l'activité de l'eau par l'utilisation de solvants organiques dans le milieu réactionnel (**figure 1**).

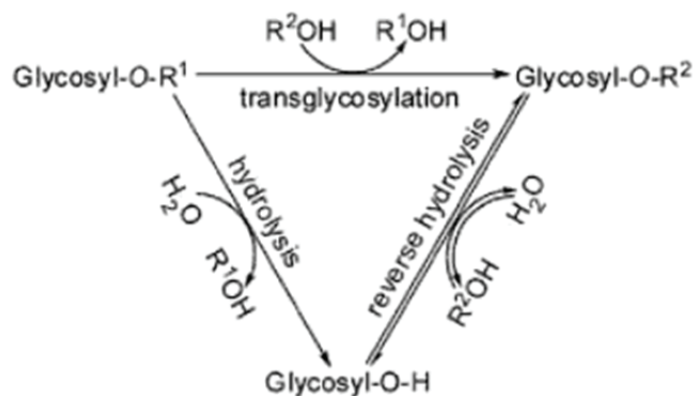
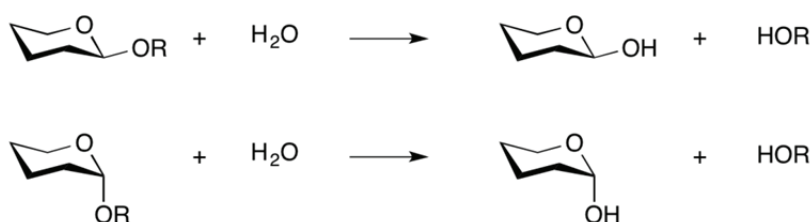


Figure 1 : Réactions catalysées par les glycoside hydrolases d'après (Rauter & Lindhorst 2011).

L'hydrolyse enzymatique des liaisons glycosidiques catalysée par les glycosides hydrolases s'effectue selon deux mécanismes principaux : le mécanisme dit « de rétention » et le mécanisme dit « d'inversion » (Koshland 1953; Chang 2012). L'hydrolyse enzymatique a généralement lieu via une catalyse nécessitant un donneur de proton ou acide, le plus souvent un acide glutamique ou un acide aspartique et un nucléophile ou base, le plus souvent un glutamate ou un aspartate. Dans les deux cas, la réaction passe par un état de transition de type ion oxocarbenium.

Le premier mécanisme est le mécanisme de rétention au cours duquel l'enzyme va en deux étapes consécutives libérer le produit d'hydrolyse avec une configuration du carbone anomérique identique à celle du substrat. Le second mécanisme est le mécanisme d'inversion au cours duquel l'enzyme va en une seule étape libérer le produit d'hydrolyse, qui possèdera une configuration du carbone anomérique qui sera opposée à celle du substrat (**figure 2**). L'appartenance d'une glycoside hydrolase à l'une de ces classes pourra être déterminée par RMN du proton en suivant la configuration du carbone anomérique du produit de la réaction.

Retaining glycoside hydrolases:



Inverting glycoside hydrolases:

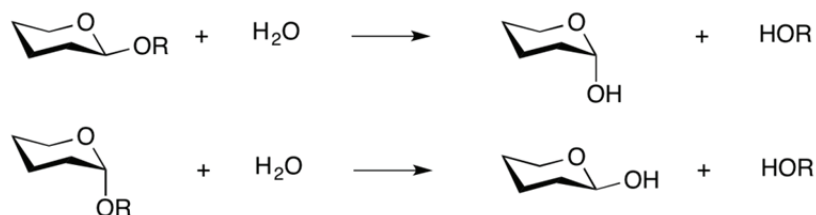


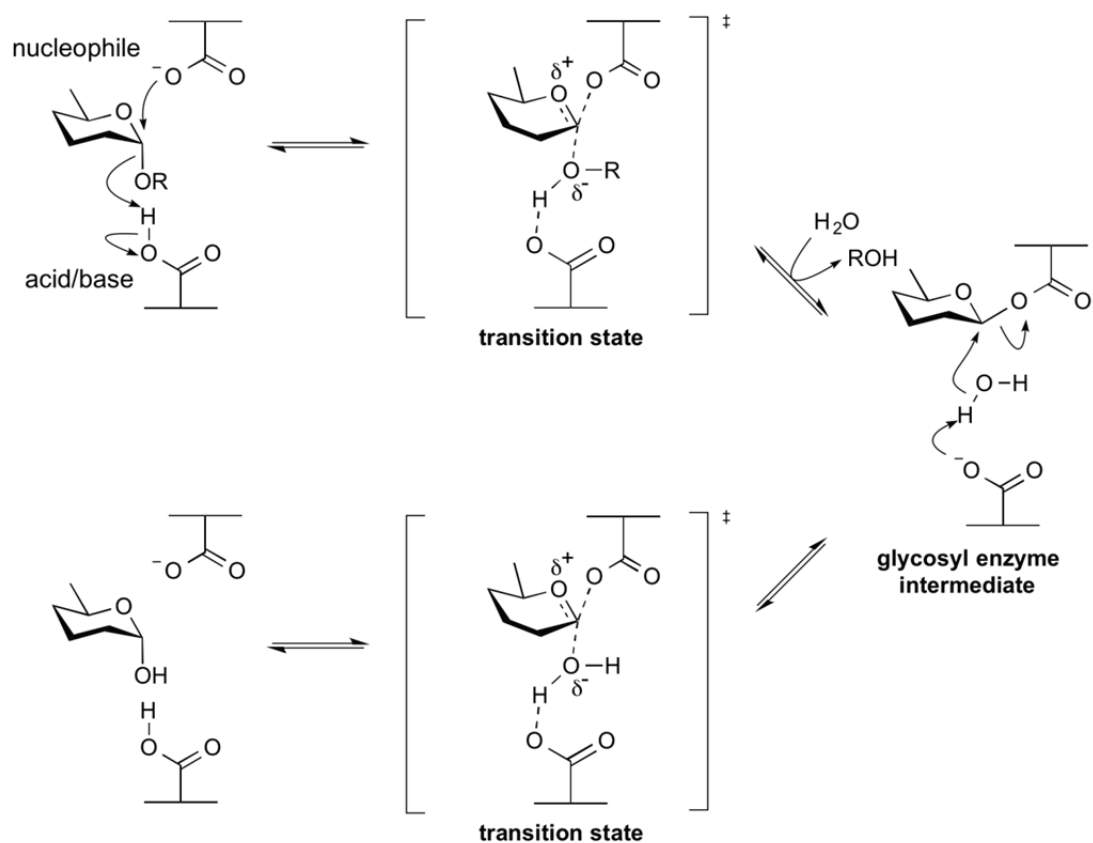
Figure 2 : Mécanismes de rétention (**haut**) et d'inversion (**bas**) pour les α et β -glycosidases d'après www.cazypedia.org.

3.1. Hydrolyse par le mécanisme de rétention

Le mécanisme d'hydrolyse avec rétention de configuration implique deux réactions d'inversion successives (**figure 3**) passant chacune par la formation d'un état de transition de type ion oxocarbénium. La première étape de l'hydrolyse (communément appelée « étape de glycosylation ») consiste en la protonation de l'atome d'oxygène de la liaison glycosidique par le résidu catalytique acide et le départ du groupement aglycone assisté par le résidu catalytique nucléophile aboutissant à la formation d'un intermédiaire glycosyl-enzyme covalent transitoire. Cette attaque concertée constitue une réaction de substitution nucléophile monomoléculaire (S_N1).

Lors de la deuxième étape (communément appelée « étape de déglycosylation »), une nouvelle substitution nucléophile intervient avec hydrolyse de l'intermédiaire covalent par une molécule d'eau jouant le rôle de nucléophile. Ceci provoque la reprotonation du résidu catalytique acide ainsi que la formation d'un produit de même conformation que le substrat. Si un autre sucre ou une molécule hydroxylée se trouvent dans le milieu alors, une réaction de transglycosylation entre en compétition avec la réaction d'hydrolyse, pouvant conduire à la formation d'une liaison glycosidique.

Retaining mechanism for an α -glycosidase:



Retaining mechanism for a β -glycosidase:

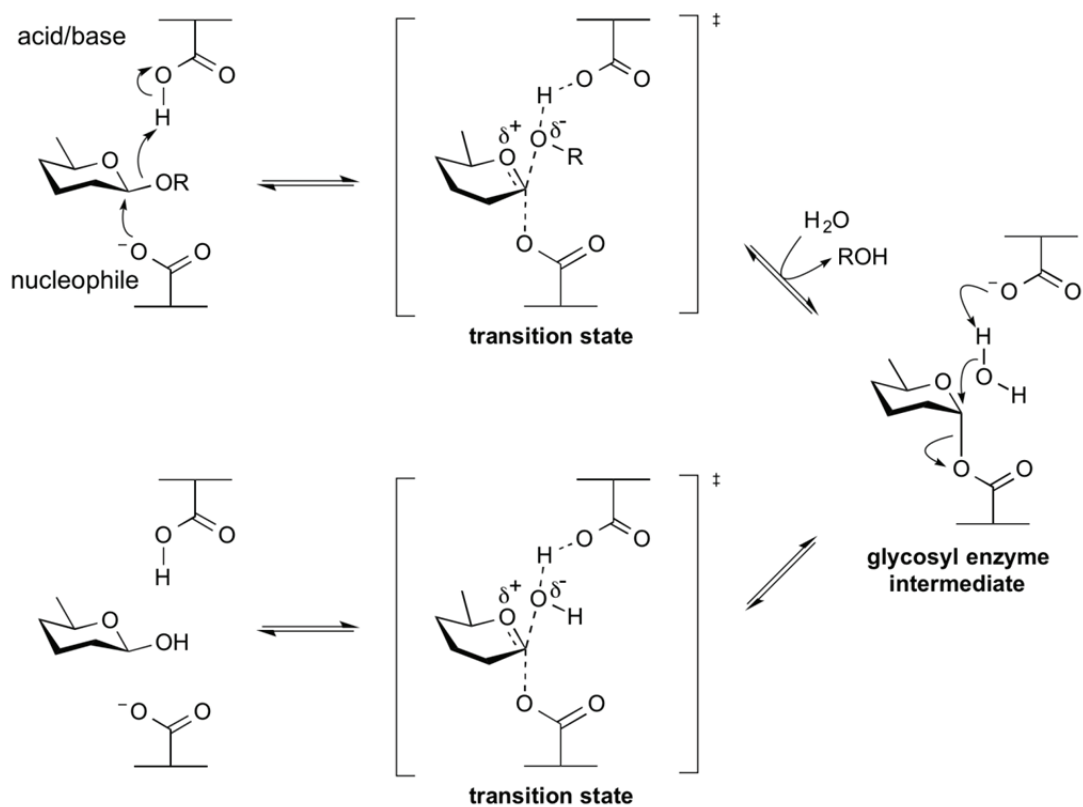
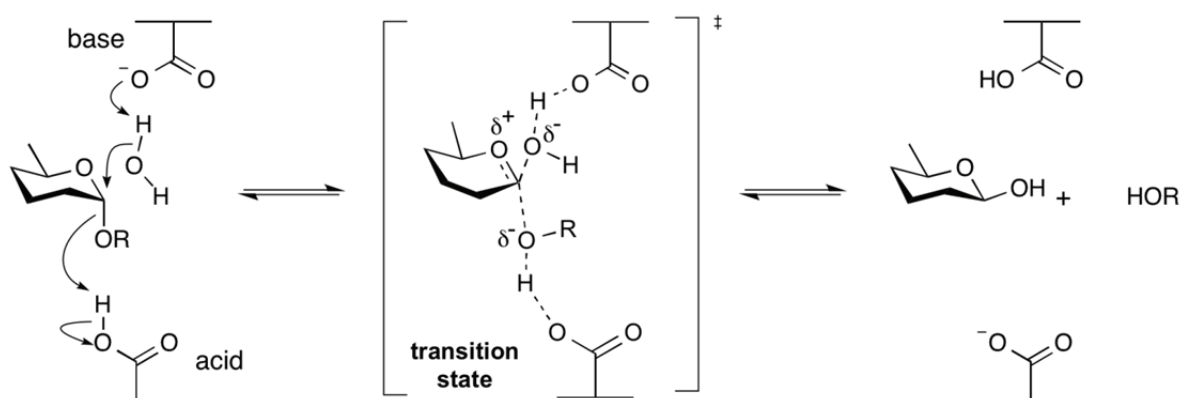


Figure 3 : Mécanisme d'hydrolyse avec rétention de la configuration du carbone anomérique pour les α (haut) et β -glycosidases (bas) d'après www.cazypedia.org.

3.2. Hydrolyse par le mécanisme d'inversion

Contrairement au mécanisme précédent, l'hydrolyse avec inversion de configuration du carbone anomérique se fait en une seule étape impliquant également la formation d'un état de transition de type ion oxocarbénium. L'oxygène de la liaison glycosidique va être protoné par le résidu catalytique acide, et le départ de l'aglycone s'accompagne de l'attaque nucléophile sur le centre anomérique d'une molécule d'eau activée par la base. Cette réaction de type SN1 va aboutir à la formation d'un produit possédant une configuration du carbone anomérique opposée à celle du substrat (**figure 4**).

Inverting mechanism for an α -glycosidase:



Inverting mechanism for a β -glycosidase:

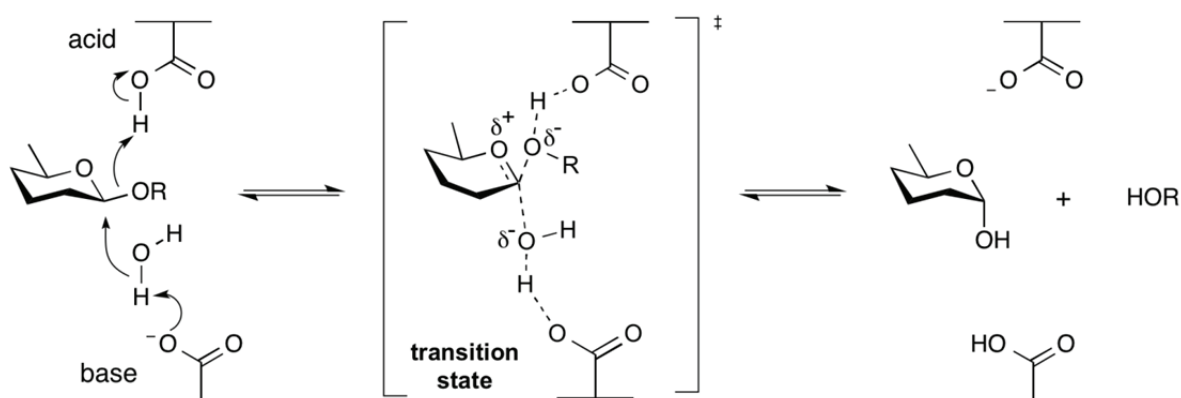


Figure 4 : Mécanisme d'hydrolyse avec inversion de la configuration du carbone anomérique pour les α (**haut**) et β -glycosides (**bas**) d'après www.cazypedia.org.

La distance séparant les résidus catalytiques est caractéristique des glycoside hydrolases fonctionnant par un mécanisme d'inversion ou par un mécanisme de rétention. Il a été observé que cette distance était de l'ordre de 5 Å pour les glycosidases fonctionnant avec un mécanisme de rétention et de 9-10 Å pour les glycosidases fonctionnant avec un mécanisme d'inversion (Vocadlo & Davies 2008).

3.3. Autres mécanismes

Les mécanismes d'hydrolyse par inversion et par rétention constituent les mécanismes catalytiques les plus observés au sein des glycoside hydrolases. Des mécanismes différents ont toutefois pu être observés chez certaines de ces enzymes (Vuong & Wilson 2010; Chang 2012).

Il a par exemple pu être observé chez certaines glycosidases l'absence d'un des deux résidus catalytiques impliquant un mécanisme alternatif faisant intervenir des nucléophiles exogènes (Hidaka *et al.* 2006), des résidus non carboxyliques (Vocadlo & Davies 2008) ou encore le substrat jouant le rôle du nucléophile (Terwisscha van Scheltinga *et al.* 1995).

4. Glycoside hydrolases et synthèse

Les glycoside hydrolases présentent, par leur activité d'hydrolyse, de nombreuses applications biotechnologiques en particulier dans le domaine agro-alimentaire, du textile ou encore des énergies renouvelables (Vuong & Wilson 2010).

La synthèse d'oligosaccharides présente également un intérêt grandissant dans l'industrie pharmaceutique de par les nombreuses applications thérapeutiques possibles de ces molécules (Perugino *et al.* 2004). La production d'oligosaccharides est difficile à mettre en œuvre par les méthodes classiques de chimie organique. En effet, la formation d'une liaison glycosidique implique un grand nombre d'étapes de protection et d'activation sélective des groupements hydroxyles, impliquant des synthèses difficiles accompagnées de faibles rendements. Une alternative intéressante consiste à utiliser des enzymes : les glycosyltransférases (EC 2.4.1) ou les glycoside hydrolases.

Les glycosyltransférases sont peu disponibles et le coût élevé de leurs substrats limitent leur exploitation. Les glycoside hydrolases, en revanche, sont plus facilement accessibles. Ces enzymes vont pouvoir catalyser la formation de liaisons glycosidiques via des réactions inverses ou via leur activité transférase.

Les glycoside hydrolases peuvent être utilisées sous leur forme sauvage dans des conditions réactionnelles favorisant la synthèse plutôt que l'hydrolyse. Le produit de la réaction reste cependant un substrat pour l'enzyme et peut être hydrolysé. Ces enzymes peuvent aussi parfois être modifiées en glycosynthase par mutagenèse dirigée (ciblant le résidu catalytique nucléophile). Dans ces conditions, la transglycosylation se fait toujours mais le produit de la réaction ne peut plus être hydrolysé (Fraser-Reid *et al.* 2001).

5. Topologie des sites actifs

Les glycoside hydrolases adoptent différents modes d'action et de reconnaissance de leurs substrats en relation avec la topologie de leur site actif. Malgré la grande diversité de structures tridimensionnelles de ces enzymes et de leurs mécanismes catalytiques, trois catégories de topologies des sites actifs peuvent être distinguées : la poche, la crevasse et le tunnel (**figure 5**).

La topologie de type poche est rencontrée chez les exo-glycosidases, et est optimale pour la reconnaissance d'un sucre de l'extrémité d'un oligo ou polysaccharide. Ces enzymes sont donc capables d'hydrolyser un oligo ou polysaccharide à partir d'un bout de chaîne en libérant un monosaccharide. Le fond de la poche accueille le sous-site donneur -1 (correspondant à l'extrémité

non réductrice du substrat) et les sous-sites +n (correspondant à l'extrémité réductrice du substrat) selon la nomenclature de G. J. Davies, le clivage ayant lieu entre le site -1 et +1 (Davies *et al.* 1997).

La topologie de type crevasse est rencontrée chez les endoglycosidases et permet la fixation de plusieurs unités saccharidiques d'un oligo ou polysaccharide. Les endo-glycosidases sont capables d'hydrolyser une chaîne oligosaccharidique de façon aléatoire.

La topologie de type tunnel est retrouvée chez les glycosidases ayant évoluées à partir d'une glycosidase à topologie crevasse, mais ayant acquis de longues boucles qui recouvrent la crevasse pour former un tunnel. Cette topologie est rencontrée chez les enzymes qui hydrolysent les polysaccharides de façon processive, et cette hydrolyse peut se faire aussi bien vers l'extrémité réductrice du substrat que vers son extrémité non réductrice selon les enzymes.

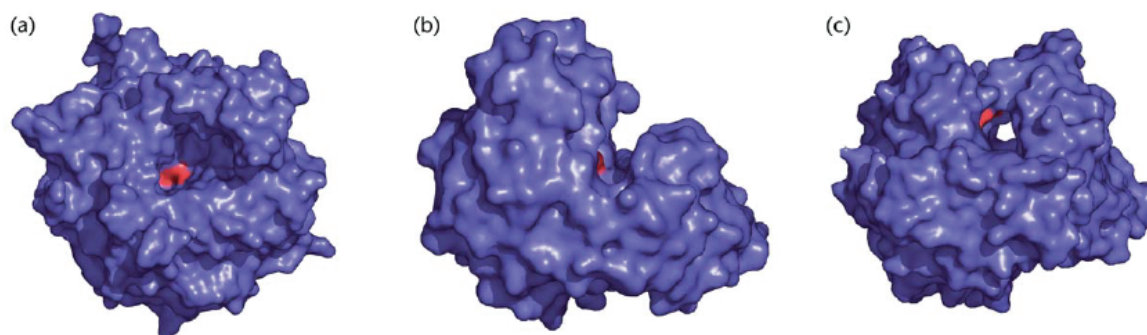


Figure 5 : Les trois topologies de sites actifs retrouvées chez les glycosidases. (A) topologie poche, (B) topologie crevasse et (C) topologie tunnel d'après (Davies & Henrissat 1995).

II. LES ALPHA-GALACTOSIDASES

1. Généralités

L' α -galactosidase ou mélibiase est une exoglycosidase qui catalyse l'hydrolyse des résidus α -galactoses non réducteurs contenus dans certains galactooligosaccharides tels que le melibiose (galactose- α -1,6-glucose), le raffinose (galactose- α -1,6-saccharose), ou le stachyose (galactose- α -1,6-raffinose) (**figures 6 et 7**). Elle est aussi capable d'hydrolyser les galactomannanes ainsi que les galactolipides et les galactoprotéines. En dehors de leur activité hydrolytique, les α -galactosidases de plantes ainsi que de microorganismes sont le plus souvent capables de catalyser des réactions de transgalactosylation ou d'hydrolyse inverse sous certaines conditions réactionnelles en particulier en présence de concentrations élevées en substrat (Eneyskaya *et al.* 1997; Zhao *et al.* 2008; Nakai *et al.* 2010).

Cette enzyme fut pour la première fois classifiée par M. A. Bau, E. Fischer et P. Lindner en 1895 en tant que mélibiase du fait de son action hydrolytique sur le disaccharide mélibiose. Plus tard, en 1928, R. Weidenhagen observa que cette enzyme est capable d'agir sur de nombreux substrats possédant des résidus α -galactoses non réducteurs et renomma la mélibiase en α -galactosidase.

Les α -galactosidases sont largement représentées dans la nature et sont présentes chez les bactéries, les champignons, les plantes ainsi que chez les animaux (Dey & Pridham 1972).

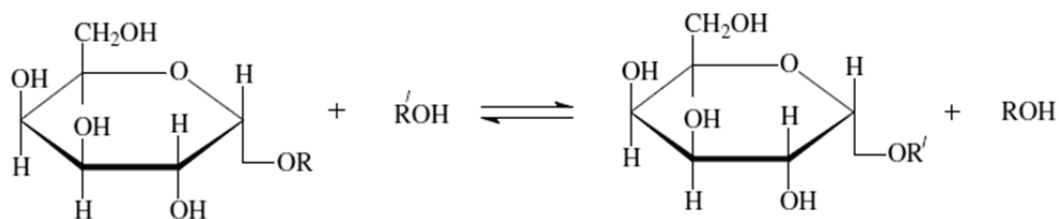


Figure 6 : Réactions catalysées par l' α -galactosidase.

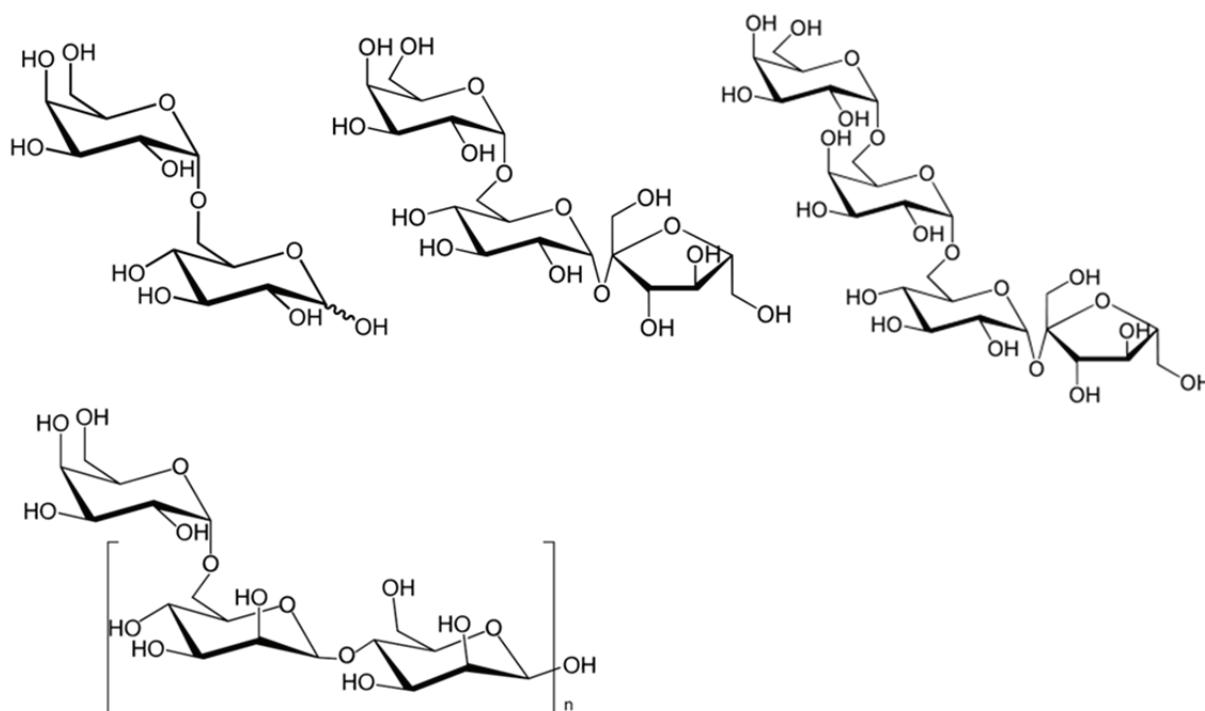


Figure 7 : Structures de galactooligosaccharides, mélibiose (α -D-Gal-(1,6)- α -D-Glu), raffinose (α -D-Gal-(1,6)- α -D-Glu-(1,2)- β -D-Fru) et stachyose (Gal-(1,6)- α -D-Gal-(1,6)- α -D-Glu-(1,2)- β -D-Fru) (**haut**) et d'un galactomannane (α -D-Gal-(1,6)-[β -D-Man-(1,4)- β -D-Man] $_n$) (**bas**).

2. Classification

Les α -galactosidases sont classifiées selon la nomenclature IUBMB et ont le numéro EC 3.2.1.22. Ces enzymes sont également classifiées selon la nomenclature de B. Henrissat en plusieurs familles GH qui sont les familles GH4, GH27, GH32, GH36, GH57, GH97 et GH110 (www.cazy.org).

La plupart des α -galactosidases appartiennent aux familles GH27 (α -galactosidases essentiellement eucaryotes) et GH36 (α -galactosidases majoritairement procaryotes). Les enzymes de la famille GH27 et GH36 sont regroupées au sein de la superfamille D ou clan GH-D avec la famille GH31 (enzymes à activités α -glucosidases et α -xylosidases) et qui sont caractérisées par un mécanisme catalytique par rétention de configuration utilisant des acides aspartiques comme résidus catalytiques et par une topologie du domaine catalytique de type tonneau (β/α)₈.

3. Sources et rôles physiologiques

Les α -galactosidases sont très représentées dans la nature et ont été purifiées à partir de nombreux organismes dont les plantes, les animaux ou encore les microorganismes.

Des α -galactosidases ont ainsi pu être isolées à partir de plantes (Dey & Pridham 1972). L'hydrolyse enzymatique des galactooligosaccharides, comme le stachyose, et des polysaccharides de la paroi cellulaire chez les végétaux, va contribuer à fournir une importante source d'énergie nécessaire à la croissance de la plante. Chez les noix de coco, il a ainsi pu être montré que ces enzymes étaient nécessaires à la germination ainsi qu'au développement de la paroi cellulaire (Balasubramaniam 1984). Il a également été reporté que l'activité de transgalactosylation des α -galactosidases pouvait également intervenir dans l'élimination des substances toxiques issues du métabolisme de la plante, appelées « phytotoxines » (Strobel 1974).

Les α -galactosidases ont également pu être isolées à partir d'animaux mais également chez l'Homme. Des α -galactosidases animales ont ainsi pu être détectées pour la première fois chez l'escargot (Bierry 1913). Ces enzymes ont également été détectées dans le cytoplasme des cellules des glandes de Brunner dans l'intestin des rats, mais également dans la moelle osseuse et les cellules sanguines chez certains animaux (Szmigielski 1966; Monis & Wasserkrug 1967). Chez l'Homme, la présence d' α -galactosidases dans la rate, le placenta, le plasma ou encore le foie ont également été rapportées (Dean & Sweeley 1979; Bishop & Desnick 1981).

Les α -galactosidases humaines sont des enzymes lysosomales qui catalysent l'hydrolyse des glycosphingolipides ainsi que des glycoprotéines. Des mutations dans le gène codant l' α -galactosidase A (gène *GLA*) peuvent résulter en l'inactivation de cette enzyme aboutissant à la maladie de Fabry, qui est une maladie récessive liée au chromosome X et touchant donc principalement les hommes avec une incidence de 1:40000 (Feldt-Rasmussen *et al.* 2002). Cette maladie s'accompagne de l'accumulation de glycosphingolipides dans de nombreux tissus et organes aboutissant le plus souvent à la mort. De même, les maladies de Schindler et Kanzaki sont deux maladies récessives autosomiques (gène localisé en 22q13) cette fois dues à un déficit en α -N-acétylgalactosaminidase et le plus souvent se déclarant chez l'enfant avec une régression neurologique et psychomotrice très rapide causant un décès entre l'âge de cinq et dix ans.

Ces enzymes ont également été retrouvées chez de nombreux microorganismes comme les levures, les champignons ou encore les bactéries (Nakai *et al.* 2010; Patil *et al.* 2010). Les α -galactosidases microbiennes peuvent dégrader des galactooligosaccharides et des polysaccharides complexes en sucres simples qui servent de sources de carbone et d'énergie. L' α -galactosidase humaine n'est pas naturellement produite dans le système digestif de l'Homme. La microflore intestinale de l'Homme est capable de dégrader les galactooligosaccharides présents dans différentes sources alimentaires s'accompagnant de la production de gaz intestinaux (Di Stefano *et al.* 2007).

4. Applications

Les activités catalytiques des α -galactosidases sont recherchées dans de nombreuses applications en particulier dans le secteur des biotechnologies et dans le secteur médical et pharmaceutique.

Ces enzymes sont par exemple utilisées dans l'industrie sucrière. La production mondiale provient à près de 30 % de la betterave sucrière. Les mélasses issues de ces betteraves contiennent de faibles quantités du galactoside raffinose. Lors du traitement de ces mélasses, la concentration en raffinose va graduellement augmenter dans le milieu aboutissant au ralentissement de la

cristallisation du sucrose ainsi qu'à la modification de sa forme cristalline, qui devient alors impropre au goût des consommateurs (Suzuki *et al.* 1969). L'élimination du raffinose dans les mélasses de betteraves par des α -galactosidases recombinantes permet d'améliorer le processus de cristallisation du sucrose ainsi que les rendements de production (le raffinose étant hydrolysé en sucrose et en galactose) (Reynolds 1974).

La plupart des mammifères dont l'Homme ne possèdent pas d' α -galactosidase pancréatique. La présence d'oligosaccharides non digestibles, comme le raffinose ou le stachyose dans certains aliments, peut ainsi induire des troubles intestinaux suite à leur consommation. Ces oligosaccharides sont dès lors fermentés par la flore intestinale aboutissant à la production de gaz intestinaux. Des suppléments diététiques en α -galactosidases recombinantes peuvent avoir des effets bénéfiques aussi bien dans l'alimentation animale que dans l'alimentation humaine (Murphy & Power 2002; Viana *et al.* 2006; Di Stefano *et al.* 2007).

L'une des applications pharmaceutiques la plus importante de ces enzymes concerne le traitement de la maladie de Fabry. L'accumulation des glycosphingolipides dans les organes liée à l'inefficacité de l' α -galactosidase mutée à hydrolyser les glycolipides et les glycoprotéines, résulte en la déficience des organes touchés et à la mort des individus. La thérapie enzymatique substitutive constitue une possibilité de traitement de cette maladie (Yoshimitsu *et al.* 2011). La complémentation chez les individus malades en α -galactosidases recombinantes s'accompagne d'une réduction effective des taux de glycosphingolipides permettant la disparition des symptômes liés à la maladie (Wilcox *et al.* 2004).

Une autre application importante des α -galactosidases concerne les transfusions sanguines. Les transfusions accidentelles d'érythrocytes non compatibles avec le sang des receveurs du fait de différences dans les antigènes sanguins ABO sont fatales pour les individus transfusés. Des méthodes enzymatiques basées sur l'utilisation d' α -galactosidases recombinantes ont été développées pour permettre la conversion d'antigènes sanguins B en antigènes sanguins O (Goldstein 1989). En effet, les antigènes sanguins B diffèrent structuralement des antigènes sanguins O uniquement par l'addition d'un résidu D-galactose terminal. L' α -galactosidase issue de graines de café s'est révélée être particulièrement efficace pour catalyser la transformation des antigènes sanguins B en antigènes sanguins O. Des cellules sanguines portant l'antigène B et traitées par cette enzyme ont ainsi pu être injectées avec succès dans le sang de patients de groupes sanguins A et O (Lenny *et al.* 1994; Zhu *et al.* 1996; Kruskall *et al.* 2000). Les α -N-acétyl galactosaminidases se sont également révélées utiles pour la transformation des antigènes sanguins B mais également A en antigènes sanguins O (Liu *et al.* 2007).

L'activité transglycosynthase de ces enzymes, qui est possible sous certaines conditions réactionnelles ou par des mutations ponctuelles de résidus catalytiques, est également recherchée (Perugino *et al.* 2004). Elle constitue une alternative intéressante à la voie de synthèse chimique pour la préparation d'oligo ou polysaccharides d'intérêt, comme les prébiotiques, qui servent de substrats sélectifs d'un nombre restreint de souches bactériennes résidant dans le système digestif et stimulent leur croissance avec des effets bénéfiques pour l'hôte.

5. Propriétés structurales

Les α -galactosidases identifiées jusqu'à présent sont retrouvées principalement au sein des familles GH27 et GH36 dans la classification de B. Henrissat. Les enzymes des familles GH27, GH36 ainsi que GH31 possèdent un mécanisme catalytique identique et une topologie générale conservée, caractérisée par la présence d'un domaine catalytique à topologie *TIM barrel* en tonneau (β/α)₈ caractéristique des enzymes du clan GH-D (Comfort *et al.* 2007).

La première structure tridimensionnelle d'un représentant du clan GH-D fut celle de la α -N-acétyl-galactosaminidase du poulet appartenant à la famille des glycoside hydrolases GH27. Cette enzyme, qui présente 50 % d'identité de séquence avec son homologue chez l'Homme, a permis de mieux comprendre les mécanismes moléculaires et structuraux de la maladie de Fabry et des maladies de Schindler et Kankazy (Garman *et al.* 2002). La structure cristalline de cette enzyme complexée avec le produit α -N-acétyl-D-galactosamine a également été résolue, permettant de définir les résidus catalytiques probablement impliqués dans le mécanisme d'hydrolyse des enzymes du clan GH-D (Garman *et al.* 2002).

Peu de temps après, les structures des α -galactosidases GH27 du riz (*Oryza sativa*), de l'Homme et d'un champignon (*Hypocrea jecorina*) seront résolues sous formes libres et complexées avec des substrats (Fujimoto *et al.* 2003; Garman & Garboczi 2004; Golubev *et al.* 2004). Ces enzymes présentent des structures quaternaires variables (rôle de certaines boucles qui ne sont pas conservées sur l'ensemble des enzymes GH27) mais possèdent des structures tertiaires très conservées.

Les structures de ces enzymes sont constituées de deux domaines. Le domaine amino-terminal constitue le domaine catalytique et présente une topologie de type *TIM barrel* caractérisée par un tonneau $(\beta/\alpha)_8$. Le second domaine ou domaine carboxy-terminal présente une topologie β -sandwich et un motif clé grecque (**figure 8**). Ce domaine est moins conservé entre les différentes α -galactosidases et sa fonction n'a pas été encore déterminée.

La résolution de plusieurs structures de l' α -galactosidase A humaine complexée avec différents composés mimant des états de transitions, permettra de préciser les détails du mécanisme moléculaire des enzymes du clan GH-D, révélant une distorsion nécessaire du substrat au cours du mécanisme catalytique (Guce *et al.* 2010).

L'étude de l' α -galactosidase tétramérique de la levure *Saccharomyces cerevisiae* permettra également de mieux comprendre les déterminants structuraux des spécificités d'oligomérisation des enzymes de la famille GH27 ainsi que les modes de reconnaissance des oligosaccharides utilisés par ces enzymes (Fernandez-Leiro *et al.* 2010).

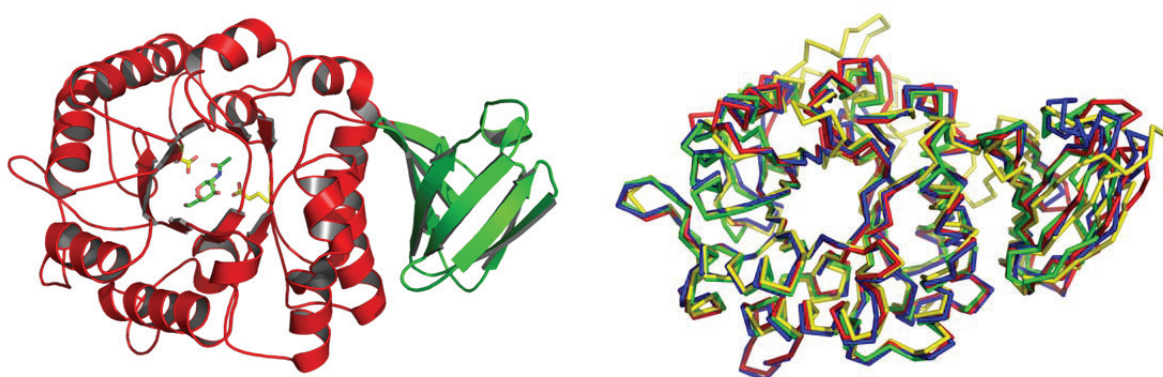


Figure 8 : (**Gauche**) structure cristalline d'un protomère de la α -N-acétyl-galactosaminidase de poulet en complexe avec le α -N-Acétyl-D-galactoside. Les résidus catalytiques sont en jaune. Le domaine catalytique est en rouge et le domaine carboxy-terminal est en vert (PDB 1KTC). (**Droite**) Superposition de différentes structures cristallines d' α -galactosidases GH27 avec celles de *Hypocrea jecorina* en jaune (PDB 1T0O), Humaine en verte (PDB 1R46), du riz en bleu (PDB 1UAS) et la α -N-acétylgalactosaminidase du poulet en rouge (PDB 1KTB).

La première structure d'une α -galactosidase appartenant à la famille GH36 sera celle de l' α -galactosidase de l'organisme hyperthermophile *Thermotoga maritima* déposée par un consortium de génomique structurale mais sans article associé (Lesley *et al.* 2002). Plus tard, la structure d'une α -galactosidase GH36 issue du microorganisme *Lactobacillus brevis* sera également résolue par un consortium de génomique structural.

Dans ces deux structures, les résidus catalytiques sont identiques à ceux observés dans les α -galactosidases GH27. On retrouve un domaine catalytique à topologie *TIM barrel* ainsi qu'un domaine carboxy-terminal à topologie β -sandwich et présentant un motif clé grecque. Toutefois, une spécificité par rapport aux α -galactosidases GH27 réside dans l'existence d'un domaine supplémentaire du côté amino-terminal et à topologie β -super sandwich. Cet extra domaine participe à la formation du site actif mais est également impliqué dans l'oligomérisation de ces enzymes (**figure 9**). Ce domaine, ainsi qu'une partie du domaine carboxy-terminal est tronqué chez l' α -galactosidase de *Thermotoga maritima* expliquant que cette enzyme soit monomérique contrairement à celle de *Lactobacillus brevis* qui est tétramérique. D'autres structures d' α -galactosidases GH36 ont récemment été résolues sous formes libres ou complexées avec le produit de la réaction α -D-galactose et présentent des structures tertiaires et quaternaires identiques à *Lactobacillus brevis* (Bruel *et al.* 2011; Fredslund *et al.* 2011).

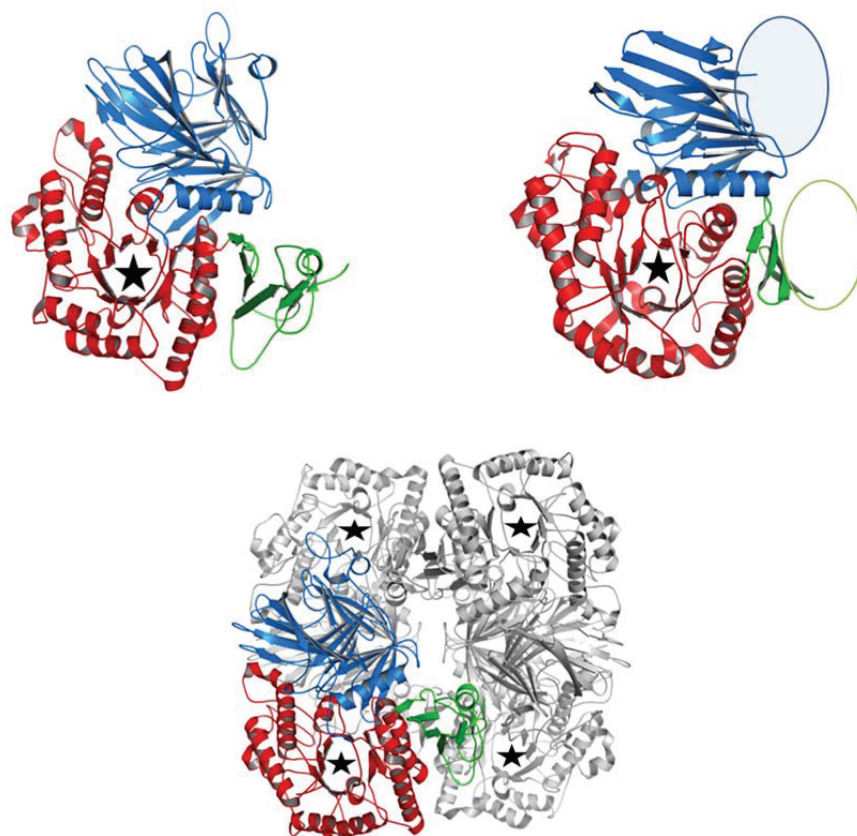


Figure 9 : (**Haut**) Structure cristalline d'un protomère de l' α -galactosidase GH36 de *Lactobacillus brevis* (PDB 3MI6) à gauche. Structure cristalline de l'enzyme de *Thermotoga maritima* (PDB 1ZY9) à droite. (**Bas**) Structure cristalline du tétramère de l' α -galactosidase GH36 de *Lactobacillus brevis*. Le site actif est repéré par une étoile noire, le domaine catalytique est en rouge, le domaine amino-terminal en bleu et le domaine carboxy-terminal en vert.

III. LES EXTRÉMOZYMES

Les enzymes sont désormais largement utilisées en tant que catalyseurs de biotransformations dans les procédés industriels du fait de leur sélectivité, versatilité et des simplifications qu'elles peuvent introduire dans des domaines encore difficiles de la chimie organique.

La plupart des enzymes utilisées jusqu'à présent est isolée à partir d'organismes mésophiles qui, par définition, ont des températures optimales de croissance entre 15 et 50 °C. Cependant, les applications industrielles de ces enzymes sont souvent limitées de par leur stabilité limitée face à des conditions environnementales extrêmes souvent retrouvées dans les applications industrielles (Hough & Danson 1999).

Les organismes extrémophiles sont des organismes isolés à partir d'environnements dits « extrêmes », où la plupart des organismes ne sont pas capables de se développer. Il existe ainsi différents phénotypes d'organismes extrémophiles en fonction de la nature des environnements d'où ils sont isolés (conditions extrêmes de température, de pression, de salinité, de pH...). La majorité des organismes extrémophiles appartient au domaine phylogénétique des archées, même si ces organismes sont également retrouvés dans le domaine des eubactéries (Brown & Doolittle 1997).

Depuis quelques années, ces organismes suscitent de plus en plus de travaux de recherche notamment à des fins industrielles. En effet, ces organismes constituent des sources d'enzymes, appelées « extrémozymes », présentant des conditions extrêmes de stabilité en comparaison avec leurs homologues mésophiles les rendant très attrayantes en tant que biocatalyseurs dans certains procédés industriels nécessitant des conditions réactionnelles extrêmes (Gomes & Steiner 2004).

Ainsi, les psychrozymes produites par les organismes dont la croissance est adaptée à des températures très basses (inférieures à 15 °C) sont intéressantes pour certaines industries, comme l'industrie du textile avec le développement des lessives ou encore l'industrie alimentaire (Gerday *et al.* 2000). Le potentiel industriel des enzymes halophiles provenant d'organismes vivant dans des milieux hypersalins est encore peu exploité, cependant, leur utilisation est en cours de développement pour le traitement des eaux usées hypersalées et l'industrie alimentaire (Oren 2010). On peut citer également le cas des thermozymes issues d'organismes se développant dans des milieux où la température est comprise entre 50 et 80 °C. L'utilisation de ces enzymes thermostables est retrouvée dans diverses industries, dont les industries de l'alimentation, chimique, pharmaceutique, du papier et de traitement des déchets (Zeikus *et al.* 1998). Un autre exemple bien connu constitue l'utilisation des hyperthermozymes provenant d'organismes se développant à des températures supérieures à 80 °C dans le domaine du génie génétique avec l'ADN polymérase provenant de l'organisme hyperthermophile *Thermus aquaticus* pour son utilisation dans la technique PCR.

Outre l'intérêt industriel des extrémozymes, les scientifiques s'intéressent également à la relation entre la structure de ces enzymes et leurs activités dans le but de comprendre comment ces extrémozymes restent fonctionnelles sous des conditions extrêmes. Il a été montré que les extrémozymes présentaient le plus souvent des structures tridimensionnelles très proches de leurs homologues non extrémophiles, et que la stabilité de ces enzymes ne semble pas liée à un seul mais à un ensemble de facteurs.

La plupart des études a été réalisée par la comparaison des thermozymes et hyperthermozymes avec leurs homologues mésophiles. Il apparaît bien souvent que les enzymes thermophiles se sont adaptées à leur environnement par des optimisations des réseaux

d'interactions au sein de leurs structures, ces dernières comprenant généralement plus de liaisons ioniques et d'autres forces internes contribuant à la stabilisation et à la rigidification de leur structure les protégeant ainsi de leur environnement. De plus, la structuration des thermozymes sous forme oligomérique et la minimisation de leur surface accessible au solvant contribuent le plus souvent à leur stabilisation. Contrairement aux thermozymes, leurs homologues mésophiles présentent en général une flexibilité structurale accrue. Cette plasticité permettrait une meilleure complémentarité du substrat avec l'enzyme, et expliquerait les hautes activités spécifiques de ces enzymes (Vieille *et al.* 1996). Les stratégies d'adaptation des extrémozymes diffèrent selon les conditions environnementales rencontrées, et peuvent être complexes. Par ailleurs, il apparaît que ces stratégies d'adaptation ne constituent pas toujours une règle générale au sein d'une même catégorie d'enzymes extrémophiles suggérant parfois des stratégies d'adaptation plus complexes.

IV. ALPHA-GALACTOSIDASES DE *GEOBACILLUS STEAROTHERMOPHILUS*

1. Généralités

C. Ganter et collaborateurs ont montré que la souche KVE39 du microorganisme thermophile *Geobacillus stearothermophilus* provenant de sources thermales islandaises, et résistante à des températures de 65 °C, possédait deux gènes *agaA* et *agaB* codant les enzymes AgaA et AgaB présentant des activités α -galactosidase. Ces gènes ont été clonés et surexprimés chez *E. coli*. La purification par précipitation à chaud des produits des gènes a permis d'obtenir les α -galactosidases AgaA et AgaB (Ganter *et al.* 1988). Ces deux isoenzymes ont une masse moléculaire de 80 kDa pour 729 acides aminés. Elles présentent 97 % d'identité de séquence soit uniquement vingt-deux acides aminés de différence sur l'ensemble de la séquence et appartiennent à la famille des glycoside hydrolases GH36 (figure 10). Des études biochimiques ont montré que ces enzymes étaient majoritairement tétramériques en solution (Fridjonsson *et al.* 1999).

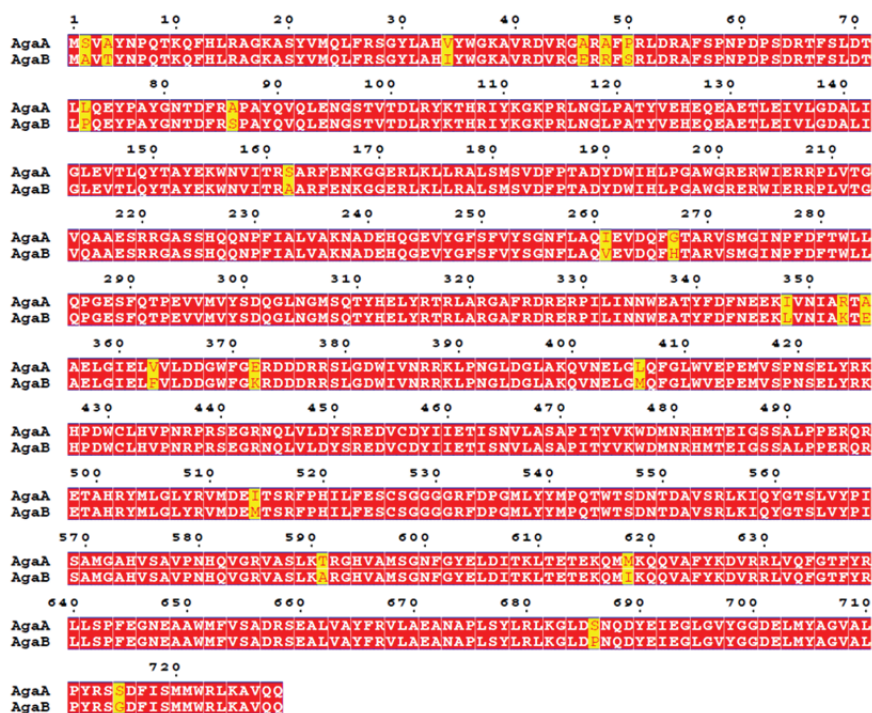


Figure 10 : Alignement de séquence protéique des enzymes AgaA et AgaB (résidus différents en jaune)

2. Propriétés catalytiques

Malgré une forte identité de séquence entre AgaA et AgaB, ces deux enzymes présentent des activités catalytiques distinctes.

En effet, les températures optimales d'activité de ces deux enzymes sont différentes (65 °C pour AgaA et 50 °C pour AgaB) malgré des stabilités thermiques très similaires (71 °C pour AgaA et 68 °C pour AgaB) (**figure 11**) (Fridjonsson *et al.* 2002). Des études biochimiques sur ces deux isoenzymes ont également permis de montrer que AgaA présentait une meilleure affinité (valeurs de K_m) pour ses substrats que AgaB, ainsi que de meilleures efficacités catalytiques (valeurs de V_m/K_m ou K_{cat}/K_m) pour la réaction d'hydrolyse (figure 11). A l'inverse, AgaB, présente une plus faible affinité pour ses substrats mais possède une bonne activité de transglycosylation. De façon intéressante, il a été montré qu'une mutation ponctuelle sur le résidu 355 (alanine chez AgaA et acide glutamique chez AgaB) permettait de transformer les propriétés catalytiques de l'enzyme AgaA en AgaB et *vice versa* (Fridjonsson *et al.* 2002).

Enzyme	AS 355	Melibiose			Raffinose		
		K_m (mM)	V_{max} (U mg ⁻¹)	V_{max}/K_m	K_m (mM)	V_{max} (U mg ⁻¹)	V_{max}/K_m
AgaA	Ala	3.3	45	13.6	6.7	48	7.2
AgaB	Glu	133	370	2.8	140	120	0.9

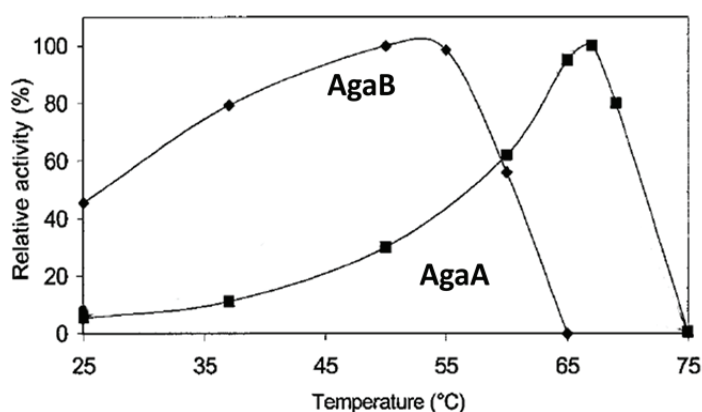


Figure 11 : (**Haut**) Paramètres cinétiques d'hydrolyse d'AgaA et AgaB. (**Bas**) Activités relatives d'AgaA et AgaB en fonction de la température d'après (Fridjonsson *et al.* 2002).

3. Applications

C. Ganter et collègues ont montré que l' α -galactosidase AgaA possédait les propriétés requises pour une application industrielle en particulier dans l'industrie sucrière. En effet, l'accumulation de D-raffinose lors du traitement des mélasses de betterave peut inhiber le processus de cristallisation du saccharose et réduire les rendements de production. L'élimination du D-raffinose par l'action d'une α -galactosidase recombinante peut permettre d'améliorer les rendements de production du saccharose. AgaA constitue une enzyme de choix car elle n'est pas inhibée par ses produits de réaction. De plus, AgaA présente une bonne affinité et une activité hydrolytique efficace contre ses substrats. Cette enzyme se trouve également stable et active sous les conditions de

manufacture du saccharose (température optimale d'activité proche de 65 °C). Enfin, le coût de cette source enzymatique est minimisé, car les souches recombinantes provenant d' *E coli* produisent l'enzyme en grande quantité (Fridjonsson *et al.* 2002).

L' α -galactosidase AgaB présente quant à elle une activité de transglycosylation qui permet des réactions d'autocondensation (Spangenberg *et al.* 2000). Cette enzyme constitue donc une alternative intéressante à la voie de synthèse chimique de disaccharides. En effet, la synthèse régiosélective de liaisons glycosidiques peut être réalisée par une approche chimique standard, mais qui nécessite des étapes de protection/déprotection difficiles à mettre en place.

OBJECTIFS DE L'ÉTUDE

Malgré une forte identité de séquence (97 %), les isoenzymes AgaA et AgaB présentent des propriétés catalytiques différentes. Il a également été montré qu'une mutation ponctuelle du résidu 355 (alanine chez AgaA et acide glutamique chez AgaB) permettait de modifier les propriétés catalytiques d'AgaB en celles d'AgaA et réciproquement. Une étude structurale a donc été initiée en collaboration avec l'équipe du Pr. R. Mattes (Institut de Génétique Industrielle, Université de Stuttgart, Allemagne) afin de mieux comprendre les déterminants moléculaires et structuraux des propriétés catalytiques différentielles de ces deux isoenzymes. Cette étude permettrait également de mieux guider le développement rationnel de nouvelles enzymes par mutagenèse dirigée, afin d'améliorer sélectivement les activités catalytiques de ces enzymes dans le sens de l'hydrolyse ou de la synthèse selon les applications recherchées. De plus, la résolution de la structure de ces enzymes contribuerait à une meilleure connaissance du mécanisme moléculaire et des propriétés structurales des α -galactosidases de la famille GH36 pour laquelle seuls quelques représentants structuraux sont connus.

Mes recherches sur ce projet se sont inscrites dans la continuité du travail de thèse de Marine Foucault (2002-2006). Ses travaux ont permis de déterminer les conditions de cristallisation des enzymes AgaB, ainsi que du mutant AgaA A335E présentant des propriétés catalytiques similaires à AgaB, et ont été publiés en 2006 (Foucault *et al.* 2006). Des données de diffraction des rayons X ont pu être collectées à 1,8 et 2,8 Å respectivement pour les cristaux d'AgaB et d'AgaA A335E. A l'époque, seule la structure cristalline de l'enzyme de *Thermotoga maritima* appartenant à la famille GH36 avait été résolue sans article associé (PDB 1ZY9). Le domaine catalytique de cette enzyme présente 24 % d'identité de séquence avec le domaine catalytique des enzymes Aga de *Geobacillus stearothermophilus*. Bien que cette identité soit faible (inférieure à 30 %), des tentatives de phasage des données par remplacement moléculaire ont été initiées, afin de remonter à l'information structurale en utilisant le modèle tridimensionnel de l'enzyme de *Thermotoga maritima* comme référence. Toutes ces tentatives n'ont cependant pas permis d'obtenir de solutions au problème de phasage. Des essais de phasage par la méthode du remplacement isomorphe multiple (méthode MIR) en utilisant des sels d'atomes lourds ont également été réalisés mais ces derniers n'ont pas permis de calculer les phases nécessaires au calcul des cartes de densité électronique et à la construction d'un modèle tridimensionnel de l'enzyme.

Mes recherches sur ce projet ont été initiées en 2011 en parallèle de mon projet intégrase. Un an auparavant, la structure de l' α -galactosidase de *Lactobacillus brevis* (PDB 3MI6) appartenant à la famille GH36 avait été résolue par un consortium de génomique structurale sans article associé. Cette enzyme, qui présente 40 % d'identité de séquence avec les enzymes Aga de *Geobacillus stearothermophilus*, a ainsi permis de résoudre la structure cristalline de AgaB et AgaA A335E par la méthode de phasage par remplacement moléculaire.

Lorsque j'ai commencé ce projet, les conditions de cristallisation de l'enzyme AgaA n'avaient pas été déterminées, les essais de cristallisation de cette enzyme n'ayant pas permis d'aboutir à la formation de cristaux analysables aux rayons X. Il était donc nécessaire de rechercher les conditions de cristallisation de cette enzyme afin d'obtenir des cristaux permettant d'obtenir la structure de l'isoenzyme AgaA pour pouvoir la comparer avec les structures des enzymes AgaB et AgaA A335E, qui présentent des propriétés catalytiques différentes de celles d'AgaA.

Une autre partie de mon travail sur ce projet a consisté à obtenir des informations structurales sur des mutants inactifs d'Aga (mutants D478A ou D548N) en complexe avec ses substrats naturels (D-raffinose et D-stachyose) ou de l'enzyme sauvage en complexe avec un inhibiteur compétitif de l'enzyme et mimant l'état de transition (1-désoxygalactonojirimycine). En effet, les structures résolues des α -galactosidases GH36 étaient alors limitées à des complexes avec le produit de la réaction d'hydrolyse, l' α -D-galactose (Bruehl *et al.* 2011; Fredslund *et al.* 2011). La résolution des structures des enzymes Aga en complexe avec des substrats naturels permettrait

d'obtenir pour la première fois des informations sur le mode de reconnaissance d'oligosaccharides naturels et substrats des α -galactosidases de la famille GH36.

Ces recherches que j'ai menées et l'article qui en a découlé sur l'analyse de ces structures (Merceron *et al.* 2012) sont décrites dans la seconde partie de ce manuscrit.

MATÉRIEL ET MÉTHODES

La majeure partie des protocoles utilisés dans cette étude est similaire à ceux de la première partie portant sur l'intégrase. Aussi, seules seront détaillées ici les parties spécifiques et qui ont été nécessaires pour la réalisation des études cristallographiques.

1. Etude structurale

1.1. Surproduction et purification des protéines

Les constructions plasmidiques des protéines sauvages ainsi que des mutants et leur surproduction en système procaryote (*E. coli*), sont réalisées par Hildegard Watzlawick de l'Université de Stuttgart, selon un protocole déjà établi (Foucault *et al.* 2006).

Brièvement, les isoenzymes AgaA, AgaB et une forme tronquée d'AgaA avec une délétion des 9 premiers résidus (AgaA-Δ9) sont surproduites et purifiées jusqu'à homogénéité. Le mutant AgaA A355E, présentant des activités catalytiques semblables à AgaB, est également surproduit et purifié.

Afin d'obtenir des informations structurales sur ces enzymes en complexe avec des substrats naturels, les surproductions et purifications des mutants catalytiques inactifs AgaA A355E D478A et AgaA A355E D548N sont également réalisées.

L'ensemble des protéines utilisées pour ce travail sont stockées dans une solution constituée d'un tampon 10 mM HEPES pH 7,0, concentrées à 5-10 mg/ml et stockées à 4 °C. La solution tampon est parfois supplémentée avec 0,02 % W/V d'azoture de sodium, afin d'améliorer la conservation des échantillons protéiques.

1.2. Essais de cristallisation

Les conditions de cristallisation des protéines AgaB (20 mM HEPES pH 6,5, 1,7 M AmSO₄), et AgaA A335E (20 % W/V PEG 3350, 0,2 M NaI pH 6,9 correspondant à la condition 10 du kit PEG/ion screen de *Hampton Research*) étaient connues lorsque j'ai débuté ce projet (Foucault *et al.* 2006). Des essais de cristallisation ont donc été initiés pour les différentes protéines de ce projet et dans ces conditions de cristallisation mais n'ont pas permis d'obtenir de cristaux.

Un criblage de nouvelles conditions de cristallisation a donc été nécessaire et réalisé à l'aide de différents kits commerciaux en ciblant principalement les conditions à base de PEG et de sulfate d'ammonium. Le robot de nanopipetage Mosquito (plateforme cristallogénèse IBCP, *TTP Labtech*) ainsi qu'un robot de visualisation RockImager 182 thermostaté à 19 °C (plateforme cristallogénèse IBCP, *Formulatrix*) sont utilisés.

Les différents kits commerciaux utilisés sont :

- PEG/ion screen (*Hampton Research*)
- PEG 2 suite (*Qiagen*)
- AmSO₄ suite (*Qiagen*)

1.3. Trempages et cryoprotection

Différents essais de trempages et de co-cristallisations ont été réalisés au cours de ce projet afin d'obtenir des informations structurales sur des complexes inhibiteurs/protéine active ou substrats naturels/protéine inactive (mutants D478A ou D548N).

Pour les essais avec la protéine active en complexe avec un inhibiteur, le composé 1-désoxygalactonojirimycine, qui est un inhibiteur compétitif connu des α -galactosidases, est utilisé (concentration maximale testée : 5 mM, *Sigma*).

Pour les essais avec la protéine inactive (mutants catalytiques) en complexe avec un substrat naturel hydrolysable, les composés D-raffinose ($\text{Gal}(\alpha 1-6)\text{Glc}(\alpha 1-2\beta)\text{Fru}$), concentration maximale testée : 150 mM, *Sigma*) et D-stachyose ($\text{Gal}(\alpha 1-6)\text{Gal}(\alpha 1-6)\text{Glc}(\alpha 1-2\beta)\text{Fru}$), concentration maximale testée : 150 mM, *Sigma*) sont également testés.

La cryoprotection des cristaux s'effectue en tout dernier lieu par l'ajout d'éthylène glycol (10 % V/V final).

RÉSULTATS ET DISCUSSION

1. Résultats obtenus

Des données de diffraction des rayons X avaient été enregistrées en 2006 (Foucault *et al.* 2006) : 1) pour AgaB jusqu'à une limite de résolution de 1,8 Å dans le groupe d'espace orthorhombique I222 et 2) pour AgaA A355E jusqu'à une limite de résolution de 2,8 Å dans le groupe d'espace trigonal P3₂21. Les tentatives de phasage par remplacement moléculaire avec la structure cristalline de l'α-galactosidase GH36 de *Thermotoga maritima* (PDB 1ZY9) furent infructueuses.

En 2010, La publication dans la PDB et sans article associé, de la structure cristalline de l'α-galactosidase GH36 de *Lactobacillus brevis* (PDB 3MI6), qui présente 40 % d'identité de séquence avec les enzymes Aga de *Geobacillus stearothermophilus*, a permis de résoudre la structure cristalline de AgaB et AgaA A355E par la méthode de phasage par remplacement moléculaire.

J'ai participé à l'affinement cristallographique de ces deux structures, puis j'ai tenté d'obtenir des cristaux de l'enzyme AgaA, qui n'avait pas pu être cristallisée. J'ai ainsi initié de nouveaux essais de cristallisation sur la protéine entière sans pouvoir obtenir de cristaux analysables aux rayons X. Une optimisation de la construction plasmidique d'AgaA a alors été réalisée par le Dr H. Watzlawick à Stuttgart, consistant en la délétion des neuf premiers résidus de la protéine (mutant AgaA-Δ9). Ces résidus sont apparus comme étant flexible dans la structure d'AgaA A355E. Cette optimisation avait ainsi pour but de réduire au maximum la flexibilité d'AgaA pour favoriser sa cristallisation. Des expériences réalisées par Dr H. Watzlawick ont par ailleurs montré que l'enzyme était toujours active sous cette forme tronquée. De nouvelles expériences de cristallogénèse ont dès lors été initiées et des cristaux ont été obtenus (conditions de cristallisation décrites dans l'article). Ces cristaux appartenant au groupe d'espace trigonal P3₂21, ont permis de résoudre la structure cristalline d'AgaA-Δ9 à une résolution de 3,2 Å.

Nous avons également tenté d'obtenir des informations structurales sur un complexe enzyme/inhibiteur compétitif. Les cristaux d'AgaB n'ont pu être reproduits et ceux de AgaA-Δ9 se sont avérés être extrêmement fragiles et présentaient un faible pouvoir de diffraction des rayons X (3,5 à 4 Å de résolution), ne nous permettant pas d'envisager aisément des expériences de trempages avec ces cristaux. J'ai donc initié des essais de cristallisation sur l'enzyme AgaA A355E dans les conditions déjà déterminées, mais sans pouvoir obtenir de nouveaux cristaux. De nouvelles conditions de cristallisation ont donc été recherchées et ont permis d'obtenir des cristaux d'AgaA A355E. L'obtention de ces cristaux appartenant au groupe d'espace trigonal P3₂21 et qui sont apparus dans de nombreuses conditions pour différentes combinaisons PEG/ions, s'est avérée être reproductible et les cristaux obtenus présentaient un pouvoir de diffraction des rayons X supérieur à AgaA (2,5 à 3 Å de résolution). Ces cristaux semblaient donc bien adaptés pour les expériences de trempages et un premier essai a pu être réalisé dans une solution contenant le composé 1-désoxygalactonojirimycine à une concentration normalement saturante pour l'enzyme (5 mM). En effet, un K_i de 0,2 μM a précédemment été déterminé par nos collaborateurs et ces derniers ont pu confirmer que ce composé était bien un inhibiteur de type compétitif pour cette enzyme. Un jeu de donnée a pu être enregistré à 2,6 Å de résolution suite à cet essai de trempage et l'analyse de la structure a révélé la présence de l'inhibiteur dans le site actif de l'enzyme.

Les structures résolues des α-galactosidases GH36 étant alors limitées à des complexes avec le produit de la réaction d'hydrolyse, l'α-D-galactose, nous avons également tenté d'obtenir des informations structurales sur l'enzyme inactive complexée avec ses substrats naturels comme le trisaccharide D-raffinose (Gal-α(1-6)-Glc-α(1-6)-Fru) ou le tétrasaccharide D-stachyose (Gal-α(1-6)-raffinose). Pour cela, une stratégie basée sur l'utilisation de mutants inactifs d'AgaA A355E (D478A ou D548N) a été employée. Les cristaux des mutants inactifs sont apparus dans des conditions identiques à ceux d'AgaA A355E et des essais de trempages ainsi que de co-cristallisations ont été réalisés dans des solutions contenant l'un ou l'autre de ces deux substrats.

Des premiers essais de trempages ont été réalisés à de faibles concentrations de substrats (5 mM). Des données de diffractions ont pu être obtenues pour ces cristaux mais ont révélé une très faible occupation des substrats dans le site actif de l'enzyme aussi bien pour le D-raffinose que le D-stachyose. Les valeurs très élevées des constantes de Michaelis (K_m) pour l'enzyme AgaA A355E envers ces substrats (centaine de mM) pouvant expliquer ces résultats, des essais de trempages et de co-cristallisation ont été réalisés avec des concentrations de substrats à leurs limites de solubilités correspondant approximativement aux valeurs de K_m , afin de tenter d'augmenter l'occupation des substrats dans le site actif de l'enzyme.

Un jeu de donnée a pu être enregistré à 2,6 Å de résolution pour le complexe AgaA A355E D548N/D-raffinose obtenu par trempage dans une solution contenant 150 mM de D-raffinose. Des essais de trempages ont également été réalisés avec une concentration élevée de D-stachyose mais ont abouti le plus souvent à une perte très rapide du pouvoir de diffraction des rayons X des cristaux. Les meilleures données ont finalement été obtenues par co-cristallisation (20 mM D-stachyose) et ont permis d'obtenir un jeu de donnée à 3,6 Å pour le complexe AgaA A355E D478A/D-stachyose. L'analyse de ces structures a révélé la présence des substrats dans le site actif des enzymes mutées.

2. Résumé de l'article et de l'analyse des structures

L'article suivant, « *The Molecular Mechanism of Thermostable α -Galactosidases AgaA and AgaB Explained by X-ray Crystallography and Mutational Studies* » présente six structures (**tableau 1 de l'article**) :

- Les structures cristallines d'AgaA- $\Delta 9$ (3,2 Å, PDB 4FNR), AgaB (1,8 Å, PDB 4FNQ) et AgaA A355E (2,8 Å, PDB 4FNP)

- La structure cristalline d'AgaA A355E en complexe avec l'inhibiteur compétitif 1-désoxygalactonojirimycine (2,6 Å, PDB 4FNS).

- Les structures cristallines d'AgaA A355E/D548N en complexe avec le substrat D-raffinose (2,6 Å, PDB 4FNT) et d'AgaA A355E/D478A en complexe avec le substrat D-stachyose (3,6 Å, PDB 4FNU).

Des enzymes tétramériques

Ces structures révèlent des protéines sous formes tétramériques représentant très probablement l'unité biologique active de ces enzymes (**figure 1A de l'article**). En effet, de précédentes études biochimiques ont mis en évidence que ces enzymes étaient tétramériques en solution (Fridjonsson *et al.* 1999).

Le tétramère est présent dans l'unité asymétrique pour les cristaux d'AgaA- $\Delta 9$ et AgaA A355E avec une symétrie ponctuelle 222 et est généré par une opération de symétrie cristallographique 222 pour le cristal d'AgaB. Il peut être décrit selon la nomenclature PQR où les axes P, Q et R correspondent respectivement aux vecteurs a, b et c de la maille cristalline dans le cristal d'AgaB.

Les protomères sont composés de trois domaines caractéristiques des α -galactosidases de la famille GH36 (**figure 1A de l'article**):

-Le domaine N-terminal (résidus 1-327) avec un repliement de type β -super-sandwich comprenant vingt brins β organisés en quatre feuillets β antiparallèles ainsi qu'une hélice α reliée au domaine catalytique

-Le domaine catalytique ou central (résidus 328-634) avec un repliement en tonneau $(\beta/\alpha)_8$ ou *TIM barrel fold* caractéristique de nombreuses glycoside hydrolases et qui contient le site actif de l'enzyme avec les résidus catalytiques D478 (résidu nucléophile) et D548 (donneur de proton) (**figure 1B de l'article**)

-Le domaine C-terminal (résidus 635-727) avec un repliement de type β sandwich avec un motif clé grecque et comprenant une hélice α et huit brins β formant deux feuillets β antiparallèles.

Les structures tertiaires et quaternaires d'AgaA- $\Delta 9$, AgaB et AgaA A355E sont très similaires et peuvent être superposées avec un rmsd de 0,5 Å sur l'ensemble du squelette carboné. Les vingt-deux résidus de différents entre AgaA et AgaB sont répartis sur l'ensemble de la chaîne polypeptidique et ne font pas partie du site actif.

Rôle du résidu 355

Le résidu 355 est une alanine dans AgaA et un acide glutamique dans AgaB et AgaA A355E. L'examen approfondi des structures résolues révèle un réarrangement structural important d'une boucle de seize résidus (résidus 335-350) entre les structures d'AgaA- $\Delta 9$, AgaB et AgaA A355E. Pour AgaB et AgaA A355E, l'acide glutamique 355, qui est localisé sur la face enfouie de la première hélice α du domaine catalytique, établit une liaison hydrogène avec l'asparagine 335 appartenant à la boucle 335-350. Cette interaction n'est pas possible dans le cas d'AgaA où le résidu 355 est une alanine (**figures 1B et 1C de l'article**).

L'analyse des structures a mis en évidence que la nature du résidu 355 avait des répercussions importantes sur le site actif de l'enzyme. Ainsi, l'interaction hydrogène observée dans AgaB et AgaA A355E entraîne un déplacement de près de 2,5 Å de la boucle 335-350. Le déplacement de cette boucle entraîne également celui du tryptophane 336 qui est très conservé et qui participe à la formation du site actif et à la fixation des substrats dans les α -galactosidases (**figure 1C de l'article**). Ce déplacement aboutit par ailleurs à un élargissement du site actif pour AgaB et AgaA A355E par rapport à AgaA- $\Delta 9$ du fait de l'éloignement du tryptophane du site actif (**figure 2 de l'article**). La position du tryptophane dans AgaA est conservée avec celle des autres structures d' α -galactosidases GH36. Le positionnement différentiel de ce résidu entre AgaB ou AgaA A355E et AgaA, a probablement des répercussions sur les activités catalytiques de ces enzymes.

Un ajustement induit nécessaire à la fixation du substrat

Nos collaborateurs ont démontré que le 1-désoxygalactonojirimycine était un inhibiteur compétitif de l'enzyme AgaA A355E avec un K_i de 0,2 μ M.

Les cristaux d'AgaA A355E trempés avec cet inhibiteur (5 mM) permettent d'observer, après analyse de la structure, la présence du ligand dans le site actif de l'enzyme (sous-site -1) et dans chaque protomère expliquant la nature compétitive de l'inhibition (**figure 3A de l'article**). Les structures de l'enzyme native AgaA A355E et en présence de l'inhibiteur sont parfaitement superposables (rmsd de 0,5 Å) et ne révèlent pas de réarrangement structural majeur suite à la fixation de ce ligand.

De façon intéressante, la fixation de l'inhibiteur dans le site actif d'AgaA A335E aboutit à la rupture de la liaison hydrogène entre l'acide glutamique 355 et l'asparagine 335, permettant un ajustement induit du site actif de l'enzyme par le ligand et un positionnement idéal du tryptophane 336 (**figure 4A de l'article**). Le positionnement du tryptophane 336 en présence du ligand est identique à celui retrouvé dans AgaA-Δ9. Cet ajustement induit nécessaire du tryptophane 336 pourrait expliquer les affinités différentielles des enzymes AgaA et AgaB ou AgaA A355E pour ces substrats. La rupture de l'interaction 355-335 en présence du ligand s'accompagne également de la création d'une poche contenant plusieurs molécules d'eau. Cette poche ne semble pas retrouvée chez AgaA et pourrait participer au mécanisme catalytique de l'enzyme (**figure 4B de l'article**).

Le ligand, qui mime l'état de transition (ion oxocarbénium), adopte une conformation chaise et établit de nombreuses liaisons hydrogènes avec la protéine au niveau du sous site -1 et sa position est conservée avec celle du produit de la réaction (α -D-galactose) dans certaines structures d' α -galactosidases connues (**figure 3A de l'article**). Ces nombreuses interactions peuvent expliquer la haute spécificité de ces enzymes pour les α -D-galactosides.

Les deux résidus catalytiques sont bien identifiés comme étant les résidus D478 et D548 et sont séparés d'environ 7 Å en adéquation avec la distance séparant les résidus catalytique pour les enzymes fonctionnant sur le mode de rétention de la configuration du carbone anomérique.

Reconnaissance des oligosaccharides

Nos collaborateurs ont montré que les enzymes AgaA A355E/D548N et AgaA A355E/D478A étaient inactives.

Les cristaux d'AgaA A355E/D548N trempés avec le D-raffinose (150 mM) et d'AgaA A355E/D478A co-cristallisés avec le D-stachyose (20 mM), permettent d'observer, après analyse des structures, la présence des ligands dans le site actif de l'enzyme (**figures 3B et 3C de l'article**). Les structures de l'enzyme native AgaA A355E et de AgaA A355E D548N ou D478A en présence des substrats sont parfaitement superposables (rmsd de 0,5 Å) et ne révèlent pas de réarrangement structural majeur suite à la fixation du ligand.

Comme pour la fixation de l'inhibiteur dans le site actif de AgaA A335E, la fixation des substrats D-raffinose et D-stachyose aboutit à la rupture de la liaison hydrogène entre l'acide glutamique 355 et l'asparagine 335, permettant un ajustement induit du site actif de l'enzyme par le substrat et un positionnement idéal du tryptophane 336 pour interagir avec le ligand.

Les interactions protéine-ligand sont nombreuses au niveau du sous site -1 et identique à celles observées dans la structure avec le 1-désoxygalactonojirimycine. Elles sont toutefois beaucoup moins nombreuses au niveau des sous sites +1 à +3, permettant probablement à l'enzyme d'être hautement spécifique des α -D-galactosides mais de pouvoir s'adapter à des α -D-galactosides avec différents sucres au niveau des sous sites +1 à +3.

Ces complexes permettent de mettre en évidence l'importance de l'oligomérisation de l'enzyme pour la fixation d'oligosaccharides. En effet, les extrémités non réductrices des sucres D-raffinose et D-stachyose interagissent avec les résidus du domaine amino-terminal d'un monomère voisin (**figure 4D de l'article**).

Importance du résidu W336 dans l'activité de l'enzyme

Des études cinétiques réalisées par nos collaborateurs ont mis en évidence que des mutations ciblant le résidu tryptophane 336, aboutissaient à des enzymes très peu efficaces pour l'activité d'hydrolyse (**tableau 2 de l'article**). Certains mutants ont présenté à l'inverse une augmentation importante de l'activité de transglycosylation. De façon intéressante, le mutant AgaA W336F présente des activités très similaires à AgaB. Ces résultats confirment bien l'importance du résidu tryptophane 336 dans les activités catalytiques de ces enzymes. Les valeurs des constantes de Michaelis-Menten obtenues pour ces différents mutants indiquent bien que le résidu tryptophane 336 est le mieux adapté pour interagir avec les substrats de ces enzymes.

L'observation des structures en complexe avec l'inhibiteur ou les substrats naturels nous a permis d'observer que le résidu W336 interagissait avec la région C4-C5-C6 de l'unité de galactose dans le sous-site -1. La fixation d'un glucose, qui est l'épimère C4 du galactose, semble impossible du fait de la position de l'hydroxyle du carbone 4 qui rentrerait en contact avec le résidu W336. Nous nous sommes donc demandés si la mutation du résidu tryptophane par des résidus plus courts pourraient permettre à l'enzyme d'hydrolyser des α -glucosides. Aucune activité α -glucosidase n'a pu être détectée pour aucun des mutants. Il apparaît donc à la vue de ce résultat que le tryptophane 336 n'est pas le seul déterminant de la spécificité de ces enzymes pour les α -galactosides.

CONCLUSION ET PERSPECTIVES

Des propriétés catalytiques différentielles

L'étude cristallographique des α -galactosidases AgaA, AgaB et AgaA A355E de la famille GH36, révèle la présence dans AgaB et AgaA A355E d'une liaison hydrogène entre l'acide glutamique 355 (E355) et l'asparagine 335 (N335). Nous avons observé que cette interaction entraînait un éloignement significatif du tryptophane 336 (W336) au niveau du sous-site catalytique -1. Cette liaison hydrogène, observée dans AgaB et AgaA A355E, est absente dans AgaA du fait de la substitution du résidu E355 par une alanine. Ainsi dans AgaA, le résidu W336, qui est adjacent au résidu N335, est engagé dans le sous-site catalytique -1 et se superpose parfaitement avec les tryptophanes équivalents retrouvés dans les autres structures d' α -galactosidases GH36 connues (Bruehl *et al.* 2011; Fredslund *et al.* 2011). De plus, aucun déplacement du tryptophane 336 n'est observé sur ces enzymes entre les formes libres ou complexées avec le produit de la réaction.

De précédentes études enzymatiques ont mis en évidence que les K_m pour les substrats naturels étaient considérablement augmentés pour l'enzyme AgaB en comparaison avec AgaA et que AgaB catalysait plus efficacement les réactions de transglycosylations (Spangenberg *et al.* 2000; Fridjonsson *et al.* 2002). La différence de positionnement du résidu W336 dans les enzymes AgaB et AgaA A355E, par rapport à AgaA, pourrait expliquer les différences d'affinités de ces enzymes pour leurs substrats avec des répercussions sur leurs efficacités catalytiques. AgaB présente des efficacités catalytiques d'hydrolyse pour ses substrats (K_{cat}/K_m) considérablement plus faibles que pour AgaA. Les produits de la transglycosylation sont ainsi plus rapidement hydrolysés pour AgaA que pour AgaB (pour des concentrations en substrat inférieures à leurs K_m chez AgaB), ceci pouvant expliquer que le bilan de la transglycosylation soit meilleur pour AgaB que pour AgaA.

Ce tryptophane est très conservé dans les familles d'enzymes GH27 et GH36 et est essentiel pour la fixation efficace de ses substrats (Fernandez-Leiro *et al.* 2010). Chez l' α -galactosidase humaine appartenant à la famille GH27, ce résidu est substitué par une glycine dans certaines formes de la maladie de Fabry (Garman *et al.* 2002). Cette maladie a pour origine un déficit enzymatique de l' α -galactosidase lysosomale avec accumulation de globotriaosylcéramide et de digalactosylcéramide dans les cellules. Ceci souligne l'importance de ce résidu dans l'activité catalytique de ces enzymes.

Le mécanisme catalytique des enzymes AgaB et AgaA A355E implique probablement la reformation de la liaison hydrogène entre les résidus N335 et E355 après le départ du produit de la réaction. L'apparition d'une poche contenant des molécules d'eau à proximité du résidu 355, suite à la fixation du substrat, pourrait contribuer au déplacement de la boucle 335-355 au cours de la réaction. Cette poche est spécifique aux enzymes AgaB et AgaA A355E et n'est pas observée chez AgaA où le résidu 355 est une alanine.

Des thermophilicités différentes

Les enzymes AgaA et AgaB sont des enzymes thermostables caractérisées par des températures de demi-dénaturation respectives de 68 et 71 °C (Fridjonsson *et al.* 2002). Malgré des stabilités thermiques similaires, les températures optimales d'activité (ou thermophilicités) d'AgaA et AgaB sont significativement différentes et respectivement de 65 et 50 °C (Fridjonsson *et al.* 2002). Par ailleurs, AgaA perd rapidement son activité avec la diminution de la température et est presque inactive à 25 °C, contrairement à AgaB, qui conserve près de 40 % de son activité maximale à cette température. Les mésozymes sont des enzymes optimalement actives entre 20 et 60 °C, alors que les thermozymes sont optimalement actives entre 60 °C et 80 °C avec une très faible activité en dessous de 40 °C (Vieille & Zeikus 2001). Ainsi, AgaA et AgaB sont des enzymes thermostables, mais AgaA peut être considérée comme étant une thermozyme alors qu'AgaB peut être considérée comme étant une mésozyme.

Des comparaisons réalisées entre les enzymes psychrophiles et leurs homologues mésophiles nous permettent de mieux comprendre ces spécificités propres aux deux enzymes. Comme mentionné précédemment, la liaison hydrogène dans AgaB et AgaA A355E entraîne un positionnement inadapté du tryptophane 336 avec pour conséquence une augmentation des constantes K_m pour les substrats naturels par rapport à AgaA. De nombreuses enzymes psychrophiles présentent des activités catalytiques (K_{cat}) améliorées à basse température au dépend des valeurs de K_m . D'un point de vue cinétique et thermodynamique, il est connu que la faible fixation d'un substrat dans le site actif d'une enzyme est avantageuse sur le plan catalytique réduisant l'énergie nécessaire pour le départ du produit (Feller & Gerday 2003; Struvay & Feller 2012). Les enzymes thermostables ont tendance à s'oligomériser de façon compacte et les enzymes Aga sont présentes sous formes tétramériques. Toutefois une exception intéressante est à noter pour l' α -galactosidase GH36 hyperthermostable de *Thermotoga maritima* (PDB 1ZY9), qui est présente sous forme monomérique du fait de l'absence de boucles impliquées dans l'oligomérisation des autres GH36 tétramériques (Comfort *et al.* 2007; Fredslund *et al.* 2011).

L'analyse des structures de psychrozymes révèle par ailleurs souvent un élargissement du site actif par rapport à leurs homologues mésophiles (Struvay & Feller 2012). C'est le cas pour AgaB et AgaA A355E où le site actif présente un élargissement du fait du positionnement du résidu W366 par rapport à AgaA. Cet élargissement du site actif a déjà été observé chez d'autres enzymes psychrophiles (Aittaleb *et al.* 1997; Russell *et al.* 1998). Ce gain en accessibilité peut là aussi participer à réduire l'énergie nécessaire pour l'accommodation du substrat dans le site actif et au départ du produit de la réaction. Ces éléments pourraient ainsi contribuer à expliquer les différences de thermophilicités observées entre AgaA et AgaB.

Un rôle dans l'évolution ?

Les enzymes AgaA et AgaB sont produites chez le microorganisme *Geobacillus stearothermophilus* et une question est soulevée quant à la pertinence biologique de ces deux isoenzymes, pour laquelle une des deux présente une activité hydrolytique fortement réduite. Il a précédemment été proposé que l'évolution des enzymes cataboliques bactériennes se faisait d'une large vers une étroite spécificité de substrat (Jensen 1976). La course à l'adaptation nécessaire pour l'acquisition de nouvelles spécificités de substrats pourrait nécessiter le développement de spécificités altérées. A un stade intermédiaire de l'évolution, certaines enzymes pourraient ainsi être retrouvées avec une large spécificité de substrats comme cela a déjà été décrit pour les enzymes pyrocatechases I et II chez *Pseudomonas* sp. B13 (Dorn & Knackmuss 1978). L'enzyme AgaB pourrait ainsi présenter une plus large spécificité de substrats par rapport à l'isoenzyme AgaA. Des analyses complémentaires pourraient permettre d'étayer cette hypothèse.

PUBLICATION 2 :

THE MOLECULAR MECHANISM OF THERMOSTABLE ALPHA-GALACTOSIDASES AGAA AND AGAB EXPLAINED BY X-RAY CRYSTALLOGRAPHY AND MUTATIONAL STUDIES

Romain Merceron, Marine Foucault, Richard Haser, Ralf Mattes, Hildegard Watzlawick,
Patrice Gouet.

J. Biol. Chem., 2012

The Molecular Mechanism of Thermostable α -Galactosidases AgaA and AgaB Explained by X-ray Crystallography and Mutational Studies^{*[S]}

Received for publication, June 28, 2012, and in revised form, September 11, 2012. Published, JBC Papers in Press, September 25, 2012, DOI 10.1074/jbc.M112.394114

Romain Merceron[‡], Marine Foucault[‡], Richard Haser[‡], Ralf Mattes[§], Hildegard Watzlawick^{§1}, and Patrice Gouet^{‡2}

From [‡]*Biocrystallography and Structural Biology of Therapeutic Targets, Bases Moléculaires et Structurales des Systèmes Infectieux-Institut de Biologie et Chimie des Protéines, UMR 5086 CNRS, Université de Lyon 1, 7 passage du Vercors, 69367 Lyon Cedex 07, France and the* [§]*Institute für Industrielle Genetik, Universität Stuttgart, Allmandring 31, 70569 Stuttgart, Germany*

Background: The thermostable α -galactosidases AgaA and AgaB are of industrial interest and have different catalytic properties despite having a high sequence identity.

Results: The crystal structures helped to explain the specificities of these two isoenzymes.

Conclusion: The A355E substitution modulates the induced fit mechanism.

Significance: Transglycosylation is increased by a single substitution that is located away from the active site.

The α -galactosidase AgaA from the thermophilic microorganism *Geobacillus stearothermophilus* has great industrial potential because it is fully active at 338 K against raffinose and can increase the yield of manufactured sucrose. AgaB has lower affinity for its natural substrates but is a powerful tool for the enzymatic synthesis of disaccharides by transglycosylation. These two enzymes have 97% identity and belong to the glycoside hydrolase (GH) family GH36 for which few structures are available. To understand the structural basis underlying the differences between these two enzymes, we determined the crystal structures of AgaA and AgaB by molecular replacement at 3.2- and 1.8 Å-resolution, respectively. We also solved a 2.8-Å structure of the AgaA_{A355E} mutant, which has enzymatic properties similar to those of AgaB. We observe that residue 355 is located 20 Å away from the active site and that the A355E substitution causes structural rearrangements resulting in a significant displacement of the invariant Trp³³⁶ at catalytic subsite –1. Hence, the active cleft of AgaA is narrowed in comparison with AgaB, and AgaA is more efficient than AgaB against its natural substrates. The structure of AgaA_{A355E} complexed with 1-deoxygalactonojirimycin reveals an induced fit movement; there is a rupture of the electrostatic interaction between Glu³⁵⁵ and Asn³³⁵ and a return of Trp³³⁶ to an optimal position for ligand stacking. The structures of two catalytic mutants of AgaA_{A355E} complexed with raffinose and stachyose show that the binding interactions are stronger at subsite –1 to enable the binding of various α -galactosides.

α -Galactosidases (α -D-galactoside galactohydrolases; EC 3.2.1.22) are widely distributed in animals, plants, and microorganisms. They catalyze the hydrolysis of terminal, non-reducing α -D-galactose residues from α -D-galactosides, including galactose oligosaccharides and polysaccharides, galactolipids, and glycoproteins. They are also known to catalyze transgalactosylation reactions at high concentrations of substrate (1). α -Galactosidases have many interesting biotechnological applications such as in animal feed processing and the pulp, paper, and sugar-producing industries where they improve the yield of sucrose by eliminating raffinose from sugar beet syrup (2–4).

Glycoside hydrolases (GHs)³ are currently divided into 130 families in the CAZy database according to homology and catalytic criteria (5). α -Galactosidases are classified into the families GH27, GH4, GH36, GH57, GH97, and GH110. Members of the families GH27, GH36, and GH31 (α -glucosidases and α -xylosidases) share a common catalytic mechanism and structural topology and form the glycoside hydrolase clan GH-D (6).

The GH27 family has been extensively studied, and numerous α -galactosidase structures from different species (bacteria and eukaryotes) have been determined (7–13). These structures are composed of two domains: the catalytic domain with a conserved (β/α)₈ barrel topology and similar active sites and the C-terminal domain. Their hydrolytic mechanism, which is common to GH36 members, proceeds with retention of the stereochemistry at the anomeric center of the substrate through a double displacement mechanism. Catalysis is mediated by two aspartate residues acting in concert as a nucleophile and a proton donor (14, 15).

Less structural information is available for the GH36 members. The crystal structures of the α -galactosidases from the hyperthermophilic microorganism *Thermotoga maritima* (Protein Data Bank code 1zy9; 564 residues) and the microorganism *Lactobacillus brevis* (Protein Data Bank code 3mi6; 745

^{*} This work was supported by the European Community's Seventh Framework Program (FP7/2007–2013) under Grant Agreement 283570 (BioStruct-X).

^[S] This article contains supplemental Fig. S1.

The atomic coordinates and structure factors (codes 4fnp, 4fnq, 4fnr, 4fns, 4fnt, and 4fnu) have been deposited in the Protein Data Bank (<http://www.pdb.org/>).

¹ To whom correspondence regarding the kinetics study may be addressed. E-mail: hildegard.watzlawick@iig.uni-stuttgart.de.

² To whom correspondence regarding the crystallographic study may be addressed. E-mail: p.gouet@ibcp.fr.

³ The abbreviations used are: GH, glycoside hydrolase; Fruf, fructofuranosyl; pNP- α Gal, para-nitrophenyl- α -galactopyranoside.

residues) have been deposited in the Protein Data Bank without accompanying publications. Those two enzymes exhibit different oligomeric states and share low sequence identity (14%). They display the same $(\beta/\alpha)_8$ barrel topology and a supplemental N-terminal domain, which is absent in the GH27 family. More recently, the crystal structures of two GH36 α -galactosidases from *Lactobacillus acidophilus* (Protein Data Bank code 2xn2; 732 residues) and *Ruminococcus gnavus* (Protein Data Bank code 2yfn; 720 residues) were determined (16, 17). They exhibit the same tetrameric organization as the *L. brevis* α -galactosidase, and the structure provides insight into the recognition of monosaccharides at the active site.

The two α -galactosidases presented herein are encoded by the genes *agaA* and *agaB* from the thermophilic microorganism *Geobacillus stearothermophilus* strain KVE39, which was isolated from Icelandic hot springs (4). AgaA and AgaB are composed of 729 amino acids each, share an identity of 97% (22 amino acids are different), and belong to the GH36 family. Despite their high sequence similarity, the AgaA and AgaB isoenzymes exhibit different catalytic properties. AgaA is of great interest for industrial applications because it is highly stable and active at 338 K (industrial processes require high temperatures) and has high affinity and hydrolytic activity against raffinose. AgaB has a lower affinity toward raffinose and reaches maximum activity at 323 K. Nevertheless, AgaB displays a better transglycosylation activity and is of interest for the enzymatic synthesis of disaccharides, which are difficult to obtain in large scale via classical organic synthesis (1, 18). Interestingly, a single mutant of AgaA, AgaA_{A355E}, exhibits catalytic properties that are similar to those of AgaB, whereas the E355A substitution in AgaB restores the catalytic properties of AgaA (19).

We solved the crystal structures of AgaA and AgaB and determined the structures of the mutant AgaA_{A355E} alone and in complex with 1-deoxygalactonojirimycin, a competitive inhibitor of α -galactosidases. The crystal structures of two catalytic mutants of AgaA_{A355E} complexed with raffinose (Gal(α 1–6)Glc(α 1–2 β)Fru_f) and stachyose (Gal(α 1–6)Gal(α 1–6)Glc(α 1–2 β)Fru_f) were also determined. Site-directed mutagenesis at position 336 of AgaA and kinetic studies show that the mutant W336F exhibits distinct enzymatic characteristics that are close to those of AgaB. These structural and molecular results explain the differences regarding activity efficiency and temperature tolerance between AgaA and AgaB. The structures also provide novel insights into substrate recognition at the sugar subunits –1 to +3 as defined in the nomenclature for sugar-binding subsites in glycoside hydrolases (20).

EXPERIMENTAL PROCEDURES

Construction, Expression, and Purification of Enzymes—The α -galactosidases AgaA, AgaA_{A355E}, and AgaB were overexpressed in *Escherichia coli* RM448 cells harboring the pBTac plasmid derivatives pAMG21, pHWG8, and pAM22 as described elsewhere (20).

The truncated form of AgaA, which lacks the first nine residues, was constructed by PCR amplification of the *agaA* gene using the oligonucleotides S7573 (gcgaattcatatgAAGCAGTTTCATTTGCGGGC) introducing an EcoRI linker and S7574

(gcctgcagTTATTGTTGAACAGCTTTC) with a PstI linker from the template plasmid pAMG21. After digestion with EcoRI and PstI, the 2178-bp fragment was inserted into pBTac1 to create the plasmid pHWG915. The active site AgaA_{A355E} mutants D478A and D548N were obtained by site-directed mutagenesis following the QuikChange® site-directed mutagenesis protocol (Stratagene). The mutations were generated using two synthetic oligonucleotides: D478A: S7746 (5'-GTGAAATGGGCTATGAACCGCCACATGACG-3') and S7747 (5'-TCATGTGGCGGGTTCATAGCCCATTTCACG-3'); D548N: S7362 (5'-AAACGTGGACTAGTAACAACACCGATGCCG-3') and S7363 (5'-GCATCGGTGTTGTTACTAGTCCGCGTTTGC-3') where the mutations are in bold face, and the silent mutations are indicated in italics. The silent mutations introduced a recognition site for the restriction enzyme SpeI in the D548N oligonucleotides and a deletion for an NdeI site in the D478A oligonucleotides. The linear amplified mutant plasmid was transformed into electrocompetent *E. coli* DH5 α cells and yielded pHWG910 (D548N) and pHWG933 (D478A). These sequences were confirmed by DNA sequencing. For expression of the genes, the plasmids were transformed in *E. coli* RM448.

Expression and purification of both native α -galactosidases and mutant enzymes followed the protocol described (20). In short, after disruption of the *E. coli* cells, the cell-free extracts were fractionated by anion-exchange chromatography on an EMD dimethylaminoethyl (DMAE) (Merck) and a Mono Q-HR 5/5 column (GE Healthcare). A final purification step was performed with a Superdex 200 column. The purified proteins were concentrated to 10 mg/ml by ultrafiltration and kept in 10 mM HEPES, pH 7.0.

Enzyme Assay—The α -galactosidase activity was determined as described (20) in 100 mM potassium phosphate buffer, pH 6.5, at 298 K. The rate of raffinose hydrolysis was determined by assessing the amount of galactose released using HPLC as described (18). Protein concentrations were determined by the Bradford method using bovine serum albumin as a standard.

Kinetic parameters (apparent Michaelis constant K_m and k_{cat}) were determined for the substrates *para*-nitrophenyl- α -galactopyranoside (pNP- α Gal) and raffinose and were obtained by curve fitting analysis using KalaidGraph (version 4.1 Synergy software). The inhibition of enzyme activity by 1-deoxygalactonojirimycin was measured at various concentrations of inhibitor (0.5–5 μ M) using pNP- α Gal as substrate. Lineweaver-Burk plots were performed for determination of the inhibition mode, and Dixon plots were used to calculate K_i . Transglycosylation reactions of Aga enzymes were performed by following the autocondensation reaction with the substrate pNP- α Gal as described (1). Products of hydrolysis (pNP) and transglycosylation (pNP- α Gal(α 1–6)Gal) were separated by thin layer chromatography and visualized under UV light at 254 nm. The amount of transglycosylation product obtained was judged qualitatively (supplemental Fig. S1 and Table 2).

Crystallization—AgaA_{A355E} and AgaB were crystallized in different space groups as described previously (21). We tried to grow new AgaB crystals but were not successful. We recently obtained crystals of a truncated form of AgaA starting at residue 10 (AgaA- Δ 9). The crystals of truncated AgaA (6.9 mg/ml

Molecular Study of AgaA and AgaB α -Galactosidases

in 10 mM HEPES, pH 7.0) were obtained using the vapor diffusion method from a solution containing 80 mM sodium acetate, 20% polyethylene glycol 4000 (w/v), and 6.4% isopropanol (w/v) at a temperature of 292 K with a 1:1 ratio of protein to reservoir solution. Although the crystallization condition was similar to that of AgaA_{A355E}, the AgaA crystals were extremely small and fragile and diffracted poorly. Only one crystal gave data to 3.2-Å resolution.

Complexes were observed for the AgaA_{A355E} type crystals, which were easily reproducible and diffracted routinely to 2.6-Å resolution. The 1-deoxygalactonojirimycin complex was obtained by soaking a crystal of AgaA_{A355E} for 2 min in a 2- μ l stabilizing solution (a reservoir solution in which the polyethylene glycol was increased to 30%) containing 5 mM inhibitor and 10% ethylene glycol (v/v) as cryoprotectant. The crystals were cooled to 100 K in liquid nitrogen before data collection.

AgaA_{A355E} is active within a large pH range (22, 23), and the crystals dissolve at extreme pH values. All attempts to obtain the crystal structures of enzyme-substrate complexes at a high or low pH where the enzyme is inactive were unsuccessful. A raffinose complex was obtained with the inactive catalytic mutant AgaA_{A355E,D548N}. The crystallization conditions were similar to those of AgaA_{A355E}, and large crystals of AgaA_{A355E,D548N} (9.2 mg/ml in 10 mM HEPES, pH 7.0) were grown from a solution containing 200 mM potassium acetate and 20% polyethylene glycol 3350 (w/v) at a pH of 8.1, a temperature of 289 K, and a 2:1 ratio of protein to reservoir solution. A crystal was soaked for 15 min in a 2- μ l stabilizing solution containing 150 mM raffinose and 10% ethylene glycol (v/v).

With the ligand stachyose, the best diffracting crystal (3.6-Å resolution) was obtained by co-crystallization with the catalytic mutant AgaA_{A355E,D478A}. This crystal (AgaA_{A355E,D478A}, 9.8 mg/ml in 10 mM HEPES, pH 7.0) was obtained from a solution containing 100 mM Tris, pH 8.5, 12% polyethylene glycol 4000 (w/v), and 20 mM stachyose at a temperature of 292 K with a 1:1 ratio. It was rapidly soaked in a stabilizing solution containing 10% (v/v) ethylene glycol and flash frozen.

X-ray Diffraction Experiments and Structure Determination—Synchrotron data were collected at 100 K on the MX beamlines at the European Synchrotron Radiation Facility in Grenoble, France and the PXIII beamline at the Swiss Light Source in Villigen, Switzerland. The programs XDS and XSCALE were used for data reduction and scaling (24). The high resolution data set collected from an AgaB crystal in 2006 (20) was recently reprocessed, and the resolution could be extended to 1.8 Å with satisfactory statistics. The phase problem was solved by molecular replacement with MOLREP (25) using *L. brevis* α -galactosidase as the search model. Five percent of the randomly selected reflections were kept apart for cross-validation, and the crystallographic refinement was performed with Phenix (26). After each cycle, the model was manually improved using the graphic program Coot (27). The structures of 1-deoxygalactonojirimycin and raffinose were extracted from the Protein Data Bank. A structure for stachyose is absent from the Protein Data Bank, and a model was generated with the program eLBOW (28). Non-crystallographic symmetry (NCS) restraints were maintained throughout the refinement in Phenix of the low resolution structure of AgaA_{A355E,D478A} complexed with

stachyose. The quality of the final crystal structures was assessed with PROCHECK (29) prior to deposition at the Protein Data Bank under the codes 4fnp (AgaA_{A355E}), 4fnq (AgaB), 4fnr (AgaA- Δ 9), 4fns (AgaA_{A355E} with 1-deoxygalactonojirimycin), 4fnt (AgaA_{A355E,D548N} with raffinose), and 4fnu (AgaA_{A355E,D478A} with stachyose). A pairwise structure comparison was accomplished with the program SUPERPOSE from the CCP4 suite (30). Images were generated with PyMOL.

RESULTS

Structure Determination of AgaA/B—The refinement statistics of the final models are presented in Table 1. The final AgaB 1.8-Å electron density map was of excellent quality, but the segment 46–49, which is located at the protein surface, was not visible. The loop 46–49, despite having high temperature factors, was visible and built in AgaA and AgaA_{A355E}. Residues 1–9, which form the first β -strand and were visible in AgaB due to crystal packing, were not observed in AgaA_{A355E}. The flexibility of the first nine residues in AgaA_{A355E} gave us the idea to express the truncated form of AgaA, which was crystallized and refined with continuous electron density maps. The regions 335–350, 371–378, 386–392, and 578–586 show high temperature factors in AgaB. These loops were better defined in AgaA and AgaA_{A355E} and did not diverge from the average temperature factor of the protein.

Although the isoforms studied are highly similar in sequence, they crystallize in different space groups. AgaA and AgaA_{A355E} crystallize in the trigonal space group P3₂21, and AgaB crystallizes in the orthorhombic space group I222 (21). The AgaA and AgaA_{A355E} crystals contain a tetramer with 222-point group symmetry in their asymmetric unit. The AgaB crystals contain one monomer in the asymmetric unit, and the tetramer is generated by a 222 crystallographic symmetry that coincides with the biological symmetry axes. Previous exclusion chromatography experiments indicate that Aga is tetrameric in solution (31) and further support the biological relevance of the observed structures. The quaternary structures of Aga can be described according to a PQR nomenclature in which P, Q, and R axes are coincident with the unit cell vectors *a*, *b*, and *c* in AgaB (Fig. 1A).

Topology—The Aga protomer is compact and folds into three domains: N-terminal, catalytic, and C-terminal domains (Fig. 1A). The N-terminal domain (residues 1–327) presents a β -supersandwich fold with 20 β -strands and four antiparallel β -sheets. It ends with a long α -helix, which is connected to the catalytic domain. One feature of interest is loop 50–66 of the P-related monomer, which is implicated in the formation of the active cleft. A second loop (residues 196–203) contributes to lining the active site. The central domain (residues 328–634) presents a (β/α)₈ triose-phosphate isomerase barrel fold (32), which is common to many glycoside hydrolases. It contains the active cleft with the fundamental catalytic residues Asp⁴⁷⁸ (nucleophile) and Asp⁵⁴⁸ (proton donor) at its bottom. The C-terminal domain (residues 635–727) presents a β -sandwich structure with a Greek key motif. It comprises one α -helix and eight β -strands that fold into two antiparallel β -sheets. It is the least conserved of the three domains. The loop 670–675 of the R-related monomer also contributes to the formation of the active cleft (Fig. 1B).

TABLE 1
Data collection and refinement statistics

Values in parentheses are for the high resolution shell. ESRF, European Synchrotron Radiation Facility; SLS, Swiss Light Source; r.m.s.d., root mean square deviation.

Data collection		AgaB	AgaA- $\Delta 9$	AgaA _{A355E}	AgaA _{A355E} + 1-deoxygalactonojirimycin	AgaA _{A355E,D548N} + raffinose	AgaA _{A355E,D178A} + stachyose
Synchrotron beamline	FIP, ESRF		ID23-2, ESRF	FIP, ESRF	ID14-4, ESRF	ID29, ESRF	PXIII, SLS
Wavelength (Å)	0.98		0.87	0.98	1.18	0.98	1.00
Space group	I222		P3 ₂ 21	P3 ₂ 21	P3 ₂ 21	P3 ₂ 21	P3 ₂ 21
Unit cell parameters (Å)	$a = 87.3, b = 113.1, c = 161.6$		$a = b = 150, c = 233$	$a = b = 150.5, c = 234.1$	$a = b = 150.3, c = 233.4$	$a = 149.7, b = 149.7, c = 234.8$	$a = 154.1, b = 154.1, c = 238.0$
Resolution limit	1.8 (1.85-1.8)		3.2 (3.28-3.2)	2.8 (3.0-2.8)	2.6 (2.67-2.6)	2.6 (2.67-2.6)	3.6 (3.69-3.6)
No. of measurements	419,569 (16,597)		219,021 (16,172)	308,452 (46,339)	196,664 (14,488)	307,727 (23,979)	208,364 (13,426)
Unique reflections	73,190 (5,287)		50,372 (3,667)	75,205 (13,526)	91,886 (6,682)	92,575 (6,859)	38,121 (2,788)
Completeness (%)	98.7 (98.1)		99.4 (100.0)	98.9 (96.2)	97.5 (96.7)	98.5 (99.7)	99.0 (99.9)
$R_{\text{sym}} (I)^a$ (%)	4.2 (26.3)		12.4 (44.9)	12.2 (43.7)	7.2 (31.1)	6.7 (38.9)	9.4 (44.7)
Mean $I/\sigma(I)$	31.08 (6.1)		12.14 (3.4)	11.28 (3.9)	11.35 (3.0)	12.22 (2.8)	18.48 (3.5)
Crystallographic refinement							
Asymmetric unit	1 monomer		1 tetramer	1 tetramer	1 tetramer	1 tetramer	1 tetramer
No. of protein atoms	5,832		23,148	23,148	23,148	23,160	23,152
No. of water molecules	396		4	71	980	357	0
Other heteroatoms	4		0	60	88	102	180
Mean B -factor (Å ²)	22		47	29.6	27.8	42.7	101.3
$R_{\text{factor}}/R_{\text{free}}^b$ (%)	18.1/20.9		17.9/22.8	19.4/24.7	17.4/23.1	17/22.1	23.8/28.2
r.m.s.d. bond lengths (Å)	0.007		0.002	0.002	0.002	0.003	0.007
r.m.s.d. bond angles (°)	1.2		0.63	0.7	0.62	0.8	1.03
Ramachandran plot favored (35)	95.1		96	96	96	96	95
Ramachandran plot outliers	(0.7)		(0.3)	(0.2)	(0.2)	(0.1)	(0.2)

^a $R_{\text{sym}} = \sum_h \sum_l |I_j - I| / \sum_h \sum_l I_j$
^b R_{factor} and R_{free} are given by $\sum |F_{\text{obs}} - F_{\text{calc}}| / \sum F_{\text{obs}}$.

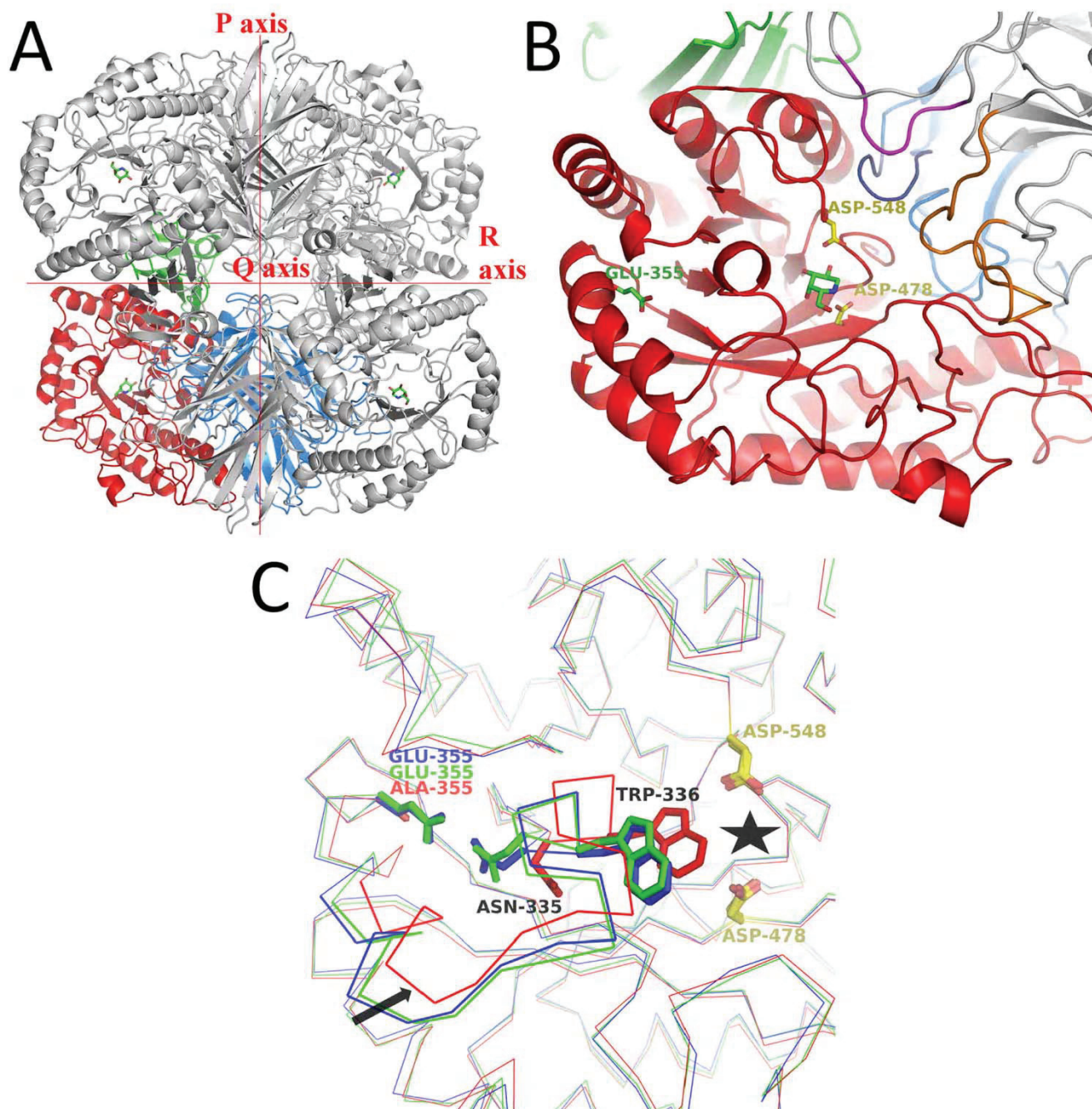


FIGURE 1. A, a schematic representation of the 222 tetrameric Aga_{A355E} in complex with 1-deoxygalactonojirimycin. The *P*, *Q*, and *R* molecular axes are shown. Monomer A is highlighted with its N-terminal domain in blue, catalytic domain in red, and C-terminal domain in green. The bound inhibitor is in green. B, a closer view of the active pocket of the liganded Aga_{A355E} with evidence of the symmetry-related monomers. The catalytic domain of monomer A is in red, and the N-terminal loop 196–203 is in deep blue. The *P*-related loop 50–66 and the *R*-related loop 670–675 are in orange and magenta, respectively. The bound 1-deoxygalactonojirimycin (green) is positioned between the catalytic Asp⁴⁷⁸ (nucleophile) and Asp⁵⁴⁸ (proton donor) in yellow. Glu³⁵⁵ (in green) is located 20 Å from the active pocket. C, superposition of AgaB, Aga_{A355E}, and AgaA and a close-up view at the 335–350 loop highlighted in thick lines. The C α traces of AgaB, Aga_{A355E}, and AgaA are in blue, green, and red, respectively. Residues 355, Trp³³⁶, and the catalytic Asp⁴⁷⁸ and Asp⁵⁴⁸ are represented as sticks. The active site is highlighted by a black star. The invariant Trp³⁶⁶ is set back from the active site cleft in AgaB and Aga_{A355E}.

AgaA Versus AgaB—AgaA and AgaB differ by 22 residues. Eleven of these residues are located in the N-terminal domain, nine are in the catalytic domain, and the remaining two are in the C-terminal domain. These 22 residues are distributed all over the polypeptide chain and are not a part of the active cleft region. A C α superposition of the AgaA and AgaB monomers reveals a significant structural rearrangement of a 16-residue segment (residues 335–350) due to residue 355, which is an

alanine in AgaA and a glutamate in AgaB. Residue 355 is located on the buried face of the first α -helix of the catalytic domain near the protein surface (Fig. 1B). In AgaB, the carboxylate of Glu³⁵⁵ is hydrogen-bonded to the extended side chain of Asn³³⁵. In AgaA, Ala³⁵⁵ makes van der Waals contacts with neighboring residues, and a bent Asn³³⁵ forms a hydrogen bond with the main-chain NH of Asn³³⁶. As a result, a significant 2.5-Å displacement of the segment 335–350 is observed

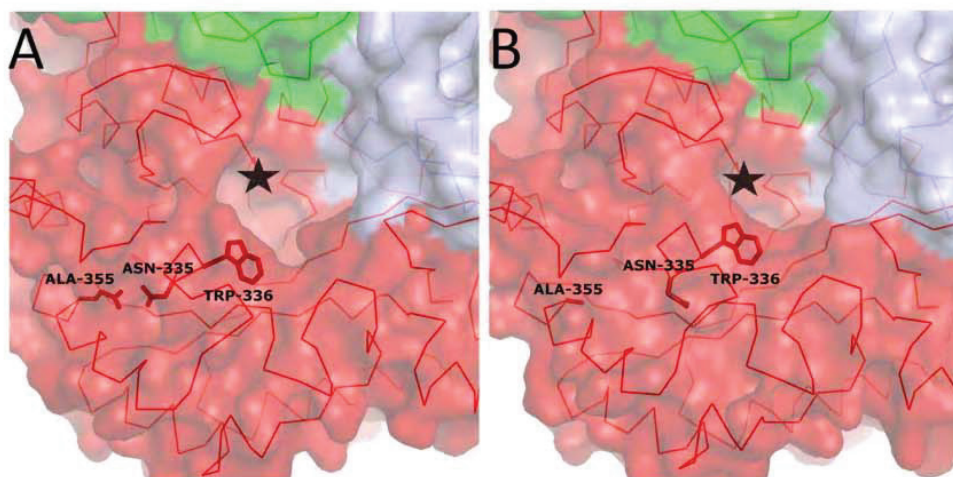


FIGURE 2. Transparent surface representations of the AgaB-like (A) and AgaA-like (B) structures are shown. The reference monomer A is colored in red. The P- and R-related monomers are in pale blue and pale green, respectively. The active cleft is highlighted by a black star. The invariant Trp³³⁶, the adjacent Asn³³⁵, and the facing Glu/Ala³⁵⁵ are highlighted. The interaction between residues Glu/Ala³⁵⁵ and Asn³³⁵ is missing in the AgaA-like structures, and the active cleft is narrowed.

between AgaA and AgaB (Fig. 1C). This movement is confirmed in the AgaA_{A355E} structure, which has a more defined electron density map. However, whereas the Asn³³⁵ side chain of AgaA_{A355E} is extended in monomers A, B, and C, it is bent in monomer D. As a consequence, the structure of the segment 335–350 in AgaA_{A355E} is quasi-identical to that of AgaB for monomers A, B, and C and to that of AgaA for monomer D. This structural heterogeneity was observed in all tested AgaA_{A355E} crystals and is probably due to crystal packing.

In addition to this segment, the tertiary and quaternary structures of AgaA, AgaB, and AgaA_{A355E} are very similar and can be superposed with a root mean square deviation of 0.5 Å on all C α pairs. For the remainder of this report, the terms “AgaA-like structure” and “AgaB-like structure” will characterize the absence and the presence of the Glu³⁵⁵–Asn³³⁵ interaction, respectively, and the accompanying shift of the segment 335–350.

The A355E Substitution—The single mutation A355E in AgaA is necessary and sufficient to transform AgaA into an AgaB-like enzyme with similar catalytic properties. Equally, the E355A mutation of AgaB restores the catalytic properties of AgaA (19). As described above, the A355E substitution induces the displacement of the segment bearing Asn³³⁵. This asparagine is adjacent to the invariant Trp³³⁶ of the active site that is involved in substrate stacking at subsite –1 (16, 17). The formation of the Glu³⁵⁵–Asn³³⁵ pair in the AgaB-like structures leads to a remote position of Trp³³⁶ at this subsite (Fig. 1C). A sequence search of GH36 α -galactosidases shows that residue 355 is a glutamate in AgaB and a serine or alanine in the remaining members. Residue 335 can be an asparagine, a serine, or a threonine. Consequently, the Glu³⁵⁵–Asn³³⁵ interaction and the movement of Trp³³⁶ at subsite –1 are probably unique to AgaB. Molecular surface representations of the tetramer show an enlargement of the active cleft in the AgaB-like structures (Fig. 2, A and B). This increase in active site accessibility has important repercussions for enzyme activity and substrate affinity and probably contributes to the different enzymatic properties of AgaA and AgaB.

AgaA_{A355E} with a Bound Inhibitor, an Induced Fit Mechanism—The inhibitory effect of 1-deoxygalactonojirimycin on AgaA_{A355E} was tested on the substrate pNP- α Gal, and a K_i of $0.2 \pm 0.03 \mu\text{M}$ was determined. From the Lineweaver-Burk plots, the inhibition mode can be considered as competitive (data not shown). AgaA_{A355E} crystals were easily reproducible and were therefore used for all of the co-crystallization and soaking experiments. Crystals of AgaA_{A355E} were soaked with the competitive inhibitor 1-deoxygalactonojirimycin. Inspection of the 2.6-Å resolution $F_o - F_c$ map reveals a bound inhibitor in the galactose binding pocket of each monomer that corresponds to subsite –1 (Fig. 3A). An AgaA-like structure with a bent Asn³³⁵ is also observed for each monomer, revealing a perfect example of an induced fit mechanism (33) with the rupture of the Glu³⁵⁵–Asn³³⁵ pair and the associated movement of the invariant Trp³³⁶ toward the ligand (Fig. 4A). A water molecule substitutes for the amide group of Asn³³⁵ and makes contact with the carboxylate of Glu³⁵⁵ when the pair is broken. Hence, a deep pocket filled with water molecules is observed near Glu³⁵⁵ (Fig. 4B). This water pocket is not observed in AgaA where residue 355 is an alanine.

Inhibitor Binding—Apart than Trp³³⁶, no significant movement is observed in subsite –1 between the native and complexed AgaA_{A355E} structures. The bound sugar ring adopts a chair conformation similar to α -galactose in the two structures of GH36 α -galactosidases complexed with reaction products (16, 17). It is located between the catalytic nucleophile Asp⁴⁷⁸ and the acid/base catalyst Asp⁵⁴⁸. These two fundamental residues are separated by 7 Å in accordance with retaining glycosidases (34). The bound inhibitor establishes numerous hydrogen bonds with surrounding protein residues (Fig. 3A). For example, N5 is firmly bonded to the carboxylate of the catalytic Asp⁴⁷⁸ and to the amide of Asn⁴⁸⁰. OH2 interacts with the main-chain NH of Gly⁵²⁹, the thiol of Cys⁵²⁶, and the catalytic Asp⁵⁴⁸. The participation of a sulfur group in ligand binding could explain the inhibitory action of Hg²⁺ and *p*-chloromercuribenzoate (22). Some GH27 and GH36 α -galactosidases cannot hydrolyze *N*-acetylgalactosides because of the presence

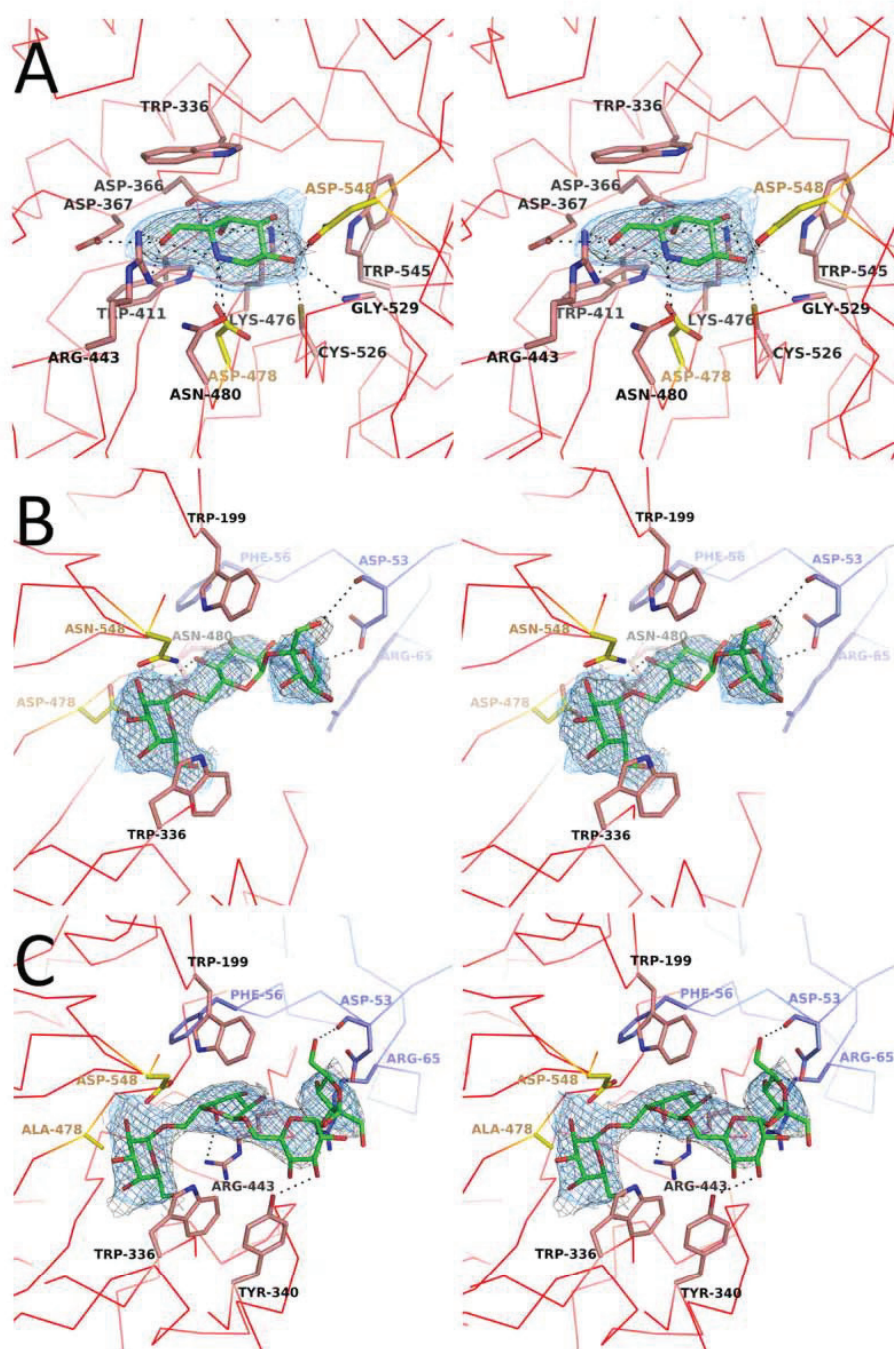


FIGURE 3. Walleye stereoview of the catalytic pocket of AgaA_{A355E} with 1-deoxygalactonojirimycin (green) (A), AgaA_{A355E,D478A} with raffinose (B), and AgaA_{A355E,D548N} with stachyose (C). Residues implicated in ligand binding are in pale red, and the mutated catalytic residues are in yellow. The $2F_o - F_c$ maps (black) are contoured at 1σ , and the $F_o - F_c$ omit maps (marine blue) calculated without ligand coordinates are contoured at 3σ . Polar contacts are represented with black dashed lines. Residues of the *P*-related monomer and those implicated in ligand binding are in blue.

of an *N*-acetyl-blocking loop at subsite -1 (7). This loop corresponds to the glycine-rich tetrapeptide Gly⁵²⁸–Gly⁵³¹ in Aga, which does not hydrolyze *N*-acetylgalactosides (31). Our liganded structure also shows that the terminal hydroxyl OH6 of the inhibitor is bound to Asp³⁶⁷ and Arg⁴⁴³. These two residues sterically prevent the existence of any subsite -2 with a C6 glycosidic linkage.

Galactose is the C4 epimer of glucose and differs only in the configuration of OH4, which is axial in galactose and equatorial in glucose. In the complex, Trp³³⁶ stacks perfectly against the C4–C5–C6 moiety of the galactose unit, whereas the axial OH4

is firmly bonded to Asp³⁶⁶, Trp⁴¹¹, and Lys⁴⁷⁶ (Fig. 3A). These structural features can explain the specificity of Aga for α -D-galactosides and the importance of the invariant Trp³³⁶ for substrate selection. The amino acids at subsite -1 are conserved in sequence among GH36 α -galactosidases, and similar patterns of protein–ligand interactions can be observed in the two structures of GH36 α -galactosidases with bound galactose (16, 17).

Raffinose and Stachyose Binding—Until now, the only known structures of α -galactosidases complexed with polysaccharides were those of the GH27 human enzyme with melibiose (15) and the GH27 *Saccharomyces cerevisiae* enzyme with melibiose and

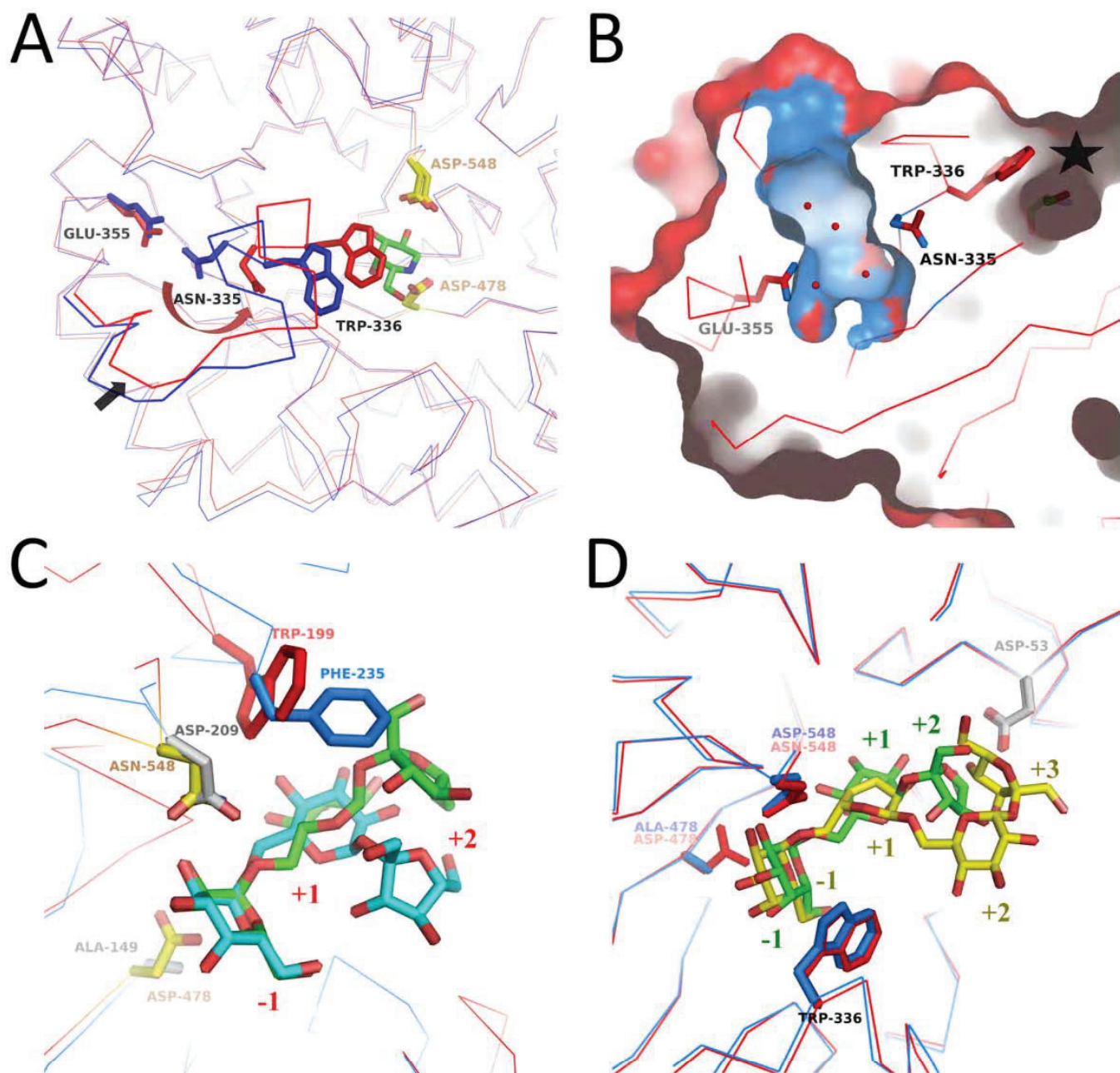


FIGURE 4. A, a close-up view of the free and liganded AgaA_{A355E} at the active -1 subsite of monomer A. The free monomer is in blue, and the complex with 1-deoxygalactonojirimycin (green) is in red. Glu³⁵⁵, Asn³³⁵, and the invariant Trp³³⁶, which stacks against the inhibitor, are shown as sticks. The catalytic residues are highlighted in yellow. The rupture of the Glu³⁵⁵-Asn³³⁵ pair upon ligand binding is represented by a red arrow, and the displacement of the 335-350 loop is highlighted by a black arrow. B, a representation of the water pocket formed upon the rupture of the Glu³⁵⁵-Asn³³⁵ pair due to the ligand binding in AgaA_{A355E}. The cavity is in blue with the water molecules. Glu³⁵⁵, Asn³³⁵, Trp³³⁶, and 1-deoxygalactonojirimycin are identified as sticks. The active site is highlighted by a black star. C, a superposition of the GH27 and GH36 active sites with bound raffinose. The C α traces of GH27 and GH36 α -galactosidases are colored in blue and red, respectively. The equivalent residues Phe²³⁵ (GH27) and Trp¹⁹⁹ (GH36) are highlighted as sticks. The mutated catalytic residues are highlighted for GH27 and GH36 in gray and yellow, respectively. Raffinose is in cyan and green for GH27 and GH36, respectively. D, a superposition of the bound raffinose (green) and stachyose (yellow) with the C α traces in red and blue, respectively. Residue Asp⁵³ of the P-related monomer is also highlighted for both structures.

raffinose (13). To better characterize the subsites -1 to +3 in GH36 α -galactosidases, the catalytic mutant AgaA_{A355E,D548N} was soaked with raffinose, and the catalytic mutant AgaA_{A355E,D478A} was co-crystallized with stachyose.

The 2.6-Å resolution electron density maps give indications of a bound raffinose in the active sites of monomers B, C, and D of AgaA_{A355E,D548N} (Fig. 3B). Monomer A is empty and retains an AgaB-like structure. Inspection of the crystal packing shows that the active cleft entrance is hindered in monomer A by

crystal contacts involving residues 337-341. The bound raffinose of GH27 *S. cerevisiae* was used as a template (13) and manually positioned in the extra density of monomers B, C, and D. The conformation of the bound raffinose at subsites +1 and +2 differs significantly between the *S. cerevisiae* enzyme (GH27) and Aga because of the replacement of Phe²³⁵ by Trp¹⁹⁹ at an equivalent position in Aga. As a result, the terminal fructose moiety is shifted by 5 Å between the two enzymes, and subsites +2 do not superimpose (Fig. 4C).

TABLE 2

AgaA/B wild-type enzymes and AgaA_{W336X} mutant kinetic data for pNP- α Gal and raffinose hydrolysis and their activity on α -glucopyranoside (α Gluc)Their transglycosylation reaction was assessed qualitatively during the autocondensation reaction with pNP- α Gal (+++, low; +++++, medium; ++++++, high; see supplemental Fig. S1). aa, amino acid; ND, not determined.

aa 336	pNP- α Gal, 25 °C			pNP- α Gluc activity ^a	Raffinose, 25 °C			Transglycosylation, visibility of autocondensation product
	K_m	k_{cat}	k_{cat}/K_m		K_m	k_{cat}	k_{cat}/K_m	
	mM	s ⁻¹	mM ⁻¹ s ⁻¹	units/mg	mM	s ⁻¹	mM ⁻¹ s ⁻¹	
AgaA WT								
Trp	0.21 \pm 0.03	105.6 \pm 7.8	502	0.000	3.8 \pm 0.6	180 \pm 17	64.3	No
Phe	0.24 \pm 0.04	375 \pm 21.4	1562	0.000	250 \pm 32	832 \pm 94	3.3	+++++
Asn	0.84 \pm 0.07	38.2 \pm 2.8	45.5	0.000	205 \pm 38	10.3 \pm 2.1	0.05	+++
Asp	1.33 \pm 0.11	0.97 \pm 0.1	0.7	0.001	ND	ND	ND	ND
Ala	0.44 \pm 0.05	7.9 \pm 0.5	17.9	0.019	252 \pm 33	48.1 \pm 6.3	0.19	+++++
Ser	8.7 \pm 2	48 \pm 7	5.5	0.000	36.6 \pm 4.2	5.2 \pm 0.3	0.14	+++++
AgaB WT								
Trp	0.83 \pm 0.09	2382 \pm 50	2870	0.000	200 \pm 27	1205 \pm 176	6.0	+++

^a Limit of detection, 0.0007 \pm 0.0002.

Extra density was observed in the 3.6-Å resolution $F_o - F_c$ map of the co-crystallized stachyose-AgaA_{A355E,D478A}, which displays a full AgaA-like structure. Hence, the problem of the accessibility of the monomer A active site was probably circumvented by the co-crystallization technique. Extra density was strong at the -1 and +1 subsites and weaker at the +2 and +3 subsites (Fig. 3C). No stachyose structure was available in the Protein Data Bank, so a model was generated and positioned into the electron density map. Although the refined stachyose fits well in the electron density, this structure has to be taken with caution because of the limited resolution of the data. Nevertheless, we can clearly observe that the bound stachyose has a pronounced kink at glucose +2, whereas the bound raffinose has an extended shape (Fig. 4D). The two oligosaccharides have a similar orientation in the cleft, and the pattern of protein-ligand interactions at subsite -1 is very similar to that of the 1-deoxygalactonojirimycin complex.

The Positive Subsites—Contacts are less extensive at subsites +1 to +3, and Aga can accommodate various α -D-galactoside substrates such as raffinose and stachyose. An opposite orientation of the glucose unit is observed at the +1 subsite between bound melibiose and bound raffinose in GH27 α -galactosidases (13). This reversal is necessary to position the subsequent sugar unit without steric clashes with the protein. For an identical reason, the +1 sugar unit of our bound raffinose and stachyose is positioned with O5 pointing to the solvent. In raffinose, the OH4 and OH3 of the +1 glucose unit is stabilized by Asn⁴⁸⁰, whereas in stachyose, the OH4 of the C4 epimer unit (galactose) is hydrogen-bonded to Arg⁴⁴³. Numerous aromatic residues are observed in the active site cleft, and the substrate contacts the bulky Trp¹⁹⁹ and the *P*-related Phe⁵⁶ at the +1 subsite.

The conformations of the bound raffinose and stachyose diverge at subsite +2. Whereas the glucose +2 of the stachyose protrudes from the sugar chain to interact with Tyr³⁴⁰, the terminal fructose of the extended raffinose is stabilized by Trp¹⁹⁹ and hydrogen-bonded to Asp⁵³ and Arg⁶⁵ of the *P*-related monomer. Moreover, the +2 subsite in raffinose overlaps with the +3 subsite in stachyose (Fig. 4D). Thus, the terminal fructose of the bound stachyose is also hydrogen-bonded to the *P*-related Asp⁵³ and Arg⁶⁵. These two residues together with Phe⁵⁶ belong to the *P*-related loop 50–66, which delineates one face of the active cleft. This loop is not conserved in the sequences of

GH36 α -galactosidases and is absent in GH27 α -galactosidases, which lack the N-terminal domain. As a consequence, the +2 subsites are not conserved between the two described GH27 and GH36 structures, and α -galactosidases may well adopt different binding strategies to accommodate long α -galactosides.

Site-directed Mutational and Kinetic Studies—We further investigated the role of AgaA Trp³³⁶ by measuring the effect of selected mutations at this position. Aliphatic, hydroxyl, polar, acidic, basic, and aromatic substitutions were tested (Ala, Ser, Asn, Asp, His, and Phe, respectively). Their hydrolysis and transglycosylation activities were measured and compared with that of the wild-type enzymes AgaA and AgaB (Table 2). Of particular interest is the AgaA_{W336F} mutant, which exhibits the highest hydrolytic efficiency against pNP- α Gal with a V_m/K_m value between that of AgaB and AgaA. The hydrolytic efficiency against raffinose was severely reduced in this mutant but still comparable with that of AgaB. The transglycosylation level was high as in AgaB. Mutations with basic residues (e.g. His or Arg) led to mostly inactive enzymes. The remaining mutants tested exhibited low hydrolysis activities. The hydrolytic decline was compensated by a good transglycosylation activity for the small residue alanine substitution. The widening of the active site probably favors the transglycosylation activity observed in AgaB. These results indicate that an aromatic/hydrophobic residue is mandatory at position 336 for the hydrolytic mechanism. The results confirm that tryptophan is best suited to stack the non-reducing galactose end (lowest K_m). The hydrolytic activities against pNP- α -glucopyranoside were also tested for all the mutants, which showed little to no activity (Table 2), demonstrating that the sole Trp³³⁶ is not responsible for the enzyme specificity toward α -D-galactosides.

DISCUSSION

Different Catalytic Properties—Our crystallographic study reveals that the presence of the Glu³⁵⁵-Asn³³⁵ pair in native AgaB involves a mispositioned Trp³³⁶ at subsite -1. This electrostatic pair is absent in AgaA because Glu³⁵⁵ is substituted by an alanine, and Trp³³⁶ stands in an optimal stacking position for the catalytic reaction. Previous enzymatic studies confirm that the K_m for natural substrates is dramatically increased (~30-fold) in AgaB in comparison with AgaA (19) and that AgaB catalyzes transglycosylation reactions in higher yields (1). The

products of transglycosylation are hydrolyzed much faster by AgaA than AgaB, which explains the apparent higher transglycosylation activity of AgaB. The A355E mutation in AgaA restores the activity properties of AgaB and vice versa, confirming the key role of the non-conserved residue 355. Equally, the mutant AgaA_{W336F} studied herein that somehow mimics the displacement of Trp³³⁶ exhibits hydrolytic and transglycosylation activities similar to those of AgaB.

Trp³³⁶ is invariant in GH27 and GH36 α -galactosidases and ensures the selection of α -galactosides by extensive stacking of the terminal non-reducing unit (13, 16, 17). In human α -galactosidase, the equivalent Trp⁴⁷ is substituted with glycine in some forms of Fabry disease (7). These results highlight the fundamental role of this tryptophan in the kinetic properties of α -galactosidases.

The catalytic mechanism of AgaB most likely involves the reformation of the Glu³⁵⁵-Asn³³⁵ pair when the reaction product leaves the active site. We observe in all our AgaB-like structures that ligand binding requires the rupture of the Glu³⁵⁵-Asn³³⁵ pair and the formation of a water pocket near Glu³⁵⁵. These water molecules in concert with the substrate could serve as a driving force that favors the displacement of the 335–350 segment during the catalytic reaction.

A significant shift in the optimum pH for the hydrolytic activity is observed between AgaA (pH 5) and AgaB or AgaA_{A355E} (pH 6.5). Again, this difference seems to be due to the A355E substitution. The possible protonation of Glu³⁵⁵ below pH 6.5 should increase the strength of the Glu³⁵⁵-Asn³³⁵ pair, now stabilized by two hydrogen bonds, and in consequence disfavor the movement of Trp³³⁶ toward subsite -1. A change in the pH can also directly affect the driving force of the water pocket, which seems to be absent in AgaA.

Different Thermophilicities—AgaA and AgaB are thermostable enzymes with temperatures of half-inactivation of 344 K for AgaA and 341 K for AgaB (19). Despite having a similar thermal stability, the thermophilicity of the two isoforms is different (*i.e.* the optimal activity temperature). Indeed, AgaA displays a temperature optimum of 338 K under the given assay conditions, whereas AgaB reaches its optimum at 323 K. AgaA loses activity quickly with decreasing temperature and is nearly inactive at 298 K, whereas AgaB retains ~40% of its maximal activity. Mesozymes are optimally active at moderate temperatures between 293 and 333 K, and thermozymes are optimally active between 333 and 353 K with low activity below 313 K (36). Consequently, AgaA can be classified as a thermozyme, but AgaB can be considered a mesozyme. Comparisons between psychrophilic and mesophilic enzymes allows us to understand the differential thermophilic properties observed between AgaA and AgaB. As mentioned previously, in AgaB-like structures, the Glu³⁵⁵-Asn³³⁵ pair induces a K_m increase for the natural substrates. Numerous psychrophilic enzymes have improved k_{cat} values at low temperatures at the expense of K_m values. From the kinetic and thermodynamic theories, it is well known that weak substrate binding is catalytically advantageous (37). Indeed, the ground state enzyme-substrate complex falls in a less deep energy pit, reducing the subsequent energy barrier that is necessary for the reaction and the energy cost at low temperatures (38).

Thermostable enzymes tend to be compact oligomers, similar to the tetrameric Aga (36). On the other hand, protein structure analyses often reveal a better accessibility to the active site in psychrophilic enzymes in comparison with their mesophilic counterparts. This is the case with the Trp³³⁶ displacement in native AgaB that widens the catalytic cleft. This larger opening of the active site was demonstrated, for example, in the structures of Antarctic fish elastase (39) and cold-active citrate synthase (40). The gain in accessibility can help to reduce the energy required for substrate accommodation and/or reaction product release in the case of macromolecular substrates. All of these elements account for the different thermophilicities of AgaA and AgaB.

An Evolutionary Role for AgaB?—The two Aga isoforms are expressed in the thermophile microorganism *G. stearothermophilus*, and a question arises regarding the physiological relevance of AgaB, which has severely impaired hydrolytic activity. It has been proposed previously that the evolution of bacterial catabolic enzymes can proceed from a broad to a narrow substrate specificity to improve metabolic efficiency (41). The course of adaptation to new substrates that are analogs to common growth substrates could require the development of altered specificities. At an intermediary stage of evolution, contemporary enzymes may be found with broad substrate specificity as was described previously for pyrocatechases I and II from *Pseudomonas* sp. B13 (42). Hence, an “adapting” AgaB can have a broader substrate activity than AgaA but with detrimental affinities.

Acknowledgments—We thank the European Synchrotron Radiation Facility and the Swiss Light Source for providing beam time and technical assistance.

REFERENCES

- Spangenberg, P., André, C., Dion, M., Rabiller, C., and Mattes, R. (2000) Comparative study of new α -galactosidases in transglycosylation reactions. *Carbohydr. Res.* **329**, 65–73
- LeBlanc, J. G., Silvestroni, A., Connes, C., Juillard, V., de Giori, G. S., Piard, J. C., and Sesma, F. (2004) Reduction of non-digestible oligosaccharides in soymilk: application of engineered lactic acid bacteria that produce α -galactosidase. *Genet. Mol. Res.* **3**, 432–440
- Clarke, J. H., Davidson, K., Rixon, J. E., Halstead, J. R., Fransen, M. P., Gilbert, H. J., and Hazlewood, G. P. (2000) A comparison of enzyme-aided bleaching of softwood paper pulp using combinations of xylanase, mannanase and α -galactosidase. *Appl. Microbiol. Biotechnol.* **53**, 661–667
- Ganter, C., Böck, A., Buckel, P., and Mattes, R. (1988) Production of thermostable, recombinant α -galactosidase suitable for raffinose elimination from sugar beet syrup. *J. Biotechnol.* **8**, 301–310
- Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* **37**, D233–D238
- Henrissat, B., and Bairoch, A. (1996) Updating the sequence-based classification of glycosyl hydrolases. *Biochem. J.* **316**, 695–696
- Garman, S. C., Hannick, L., Zhu, A., and Garboczi, D. N. (2002) The 1.9 Å structure of α -N-acetylgalactosaminidase: molecular basis of glycosidase deficiency diseases. *Structure* **10**, 425–434
- Fujimoto, Z., Kaneko, S., Momma, M., Kobayashi, H., and Mizuno, H. (2003) Crystal structure of rice α -galactosidase complexed with D-galactose. *J. Biol. Chem.* **278**, 20313–20318
- Garman, S. C., and Garboczi, D. N. (2004) The molecular defect leading to

- Fabry disease: structure of human α -galactosidase. *J. Mol. Biol.* **337**, 319–335
10. Golubev, A. M., Nagem, R. A., Brandão Neto, J. R., Neustroev, K. N., Eneyskaya, E. V., Kulminkaya, A. A., Shabalin, K. A., Savel'ev, A. N., and Polikarpov, I. (2004) Crystal structure of α -galactosidase from *Trichoderma reesei* and its complex with galactose: implications for catalytic mechanism. *J. Mol. Biol.* **339**, 413–422
11. Clark, N. E., and Garman, S. C. (2009) The 1.9 Å structure of human α -N-acetylgalactosaminidase: the molecular basis of Schindler and Kan-zaki diseases. *J. Mol. Biol.* **393**, 435–447
12. Fujimoto, Z., Kaneko, S., Kim, W. D., Park, G. G., Momma, M., and Kobayashi, H. (2009) The tetramer structure of the glycoside hydrolase family 27 α -galactosidase I from *Umbelopsis vinacea*. *Biosci. Biotechnol. Biochem.* **73**, 2360–2364
13. Fernández-Leiro, R., Pereira-Rodríguez, A., Cerdán, M. E., Becerra, M., and Sanz-Aparicio, J. (2010) Structural analysis of *Saccharomyces cerevisiae* α -galactosidase and its complexes with natural substrates reveals new insights into substrate specificity of GH27 glycosidases. *J. Biol. Chem.* **285**, 28020–28033
14. Comfort, D. A., Bobrov, K. S., Ivanen, D. R., Shabalin, K. A., Harris, J. M., Kulminkaya, A. A., Brumer, H., and Kelly, R. M. (2007) Biochemical analysis of *Thermotoga maritima* GH36 α -galactosidase (TmGalA) confirms the mechanistic commonality of clan GH-D glycoside hydrolases. *Biochemistry* **46**, 3319–3330
15. Guce, A. I., Clark, N. E., Salgado, E. N., Ivanen, D. R., Kulminkaya, A. A., Brumer, H., 3rd, and Garman, S. C. (2010) Catalytic mechanism of human α -galactosidase. *J. Biol. Chem.* **285**, 3625–3632
16. Fredslund, F., Hachem, M. A., Larsen, R. J., Sørensen, P. G., Coutinho, P. M., Lo Leggio, L., and Svensson, B. (2011) Crystal structure of α -galactosidase from *Lactobacillus acidophilus* NCFM: insight into tetramer formation and substrate binding. *J. Mol. Biol.* **412**, 466–480
17. Bruel, L., Sulzenbacher, G., Cervera Tison, M., Pujol, A., Nicoletti, C., Perrier, J., Galinier, A., Ropartz, D., Fons, M., Pompeo, F., and Giardina, T. (2011) α -Galactosidase/sucrose kinase (AgaSK), a novel bifunctional enzyme from the human microbiome coupling galactosidase and kinase activities. *J. Biol. Chem.* **286**, 40814–40823
18. Perugino, G., Trincone, A., Rossi, M., and Moracci, M. (2004) Oligosaccharide synthesis by glycosynthases. *Trends Biotechnol.* **22**, 31–37
19. Fridjonsson, O., Watzlawick, H., and Mattes, R. (2002) Thermoadaptation of α -galactosidase AgaB1 in *Thermus thermophilus*. *J. Bacteriol.* **184**, 3385–3391
20. Davies, G. J., Wilson, K. S., and Henrissat, B. (1997) Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem. J.* **321**, 557–559
21. Foucault, M., Watzlawick, H., Mattes, R., Haser, R., and Gouet, P. (2006) Crystallization and preliminary x-ray diffraction studies of two thermostable α -galactosidases from glycoside hydrolase family 36. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **62**, 100–103
22. Gote, M. M., Khan, M. I., Gokhale, D. V., Bastawde, K. B., and Khire, J. M. (2006) Purification, characterization and substrate specificity of thermostable α -galactosidase from *Bacillus stearothermophilus* (NCIM-5146). *Process Biochem.* **41**, 1311–1317
23. Gote, M. M., Khan, M. I., and Khire, J. M. (2007) Active site directed chemical modification of α -galactosidase from *Bacillus stearothermophilus* (NCIM 5146): involvement of lysine, tryptophan and carboxylate residues in catalytic site. *Enzyme Microb. Technol.* **40**, 1312–1320
24. Kabsch, W. (2010) Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 133–144
25. Vagin, A., and Teplyakov, A. (2010) Molecular replacement with MOL-REP. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 22–25
26. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Echols, N., Headd, J. J., Hung, L. W., Jain, S., Kapral, G. J., Grosse Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2011) The Phenix software for automated determination of macromolecular structures. *Methods* **55**, 94–106
27. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501
28. Moriarty, N. W., Grosse-Kunstleve, R. W., and Adams, P. D. (2009) Electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 1074–1080
29. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291
30. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242
31. Fridjonsson, O., Watzlawick, H., Gehweiler, A., and Mattes, R. (1999) Thermostable α -galactosidase from *Bacillus stearothermophilus* NUB3621: cloning, sequencing and characterization. *FEMS Microbiol. Lett.* **176**, 147–153
32. Banner, D. W., Bloomer, A. C., Petsko, G. A., Phillips, D. C., Pogson, C. I., Wilson, I. A., Corran, P. H., Furth, A. J., Milman, J. D., Offord, R. E., Priddle, J. D., and Waley, S. G. (1975) Structure of chicken muscle triose phosphate isomerase determined crystallographically at 2.5 angstrom resolution using amino acid sequence data. *Nature* **255**, 609–614
33. Koshland, D. E. (1958) Application of a theory of enzyme specificity to protein synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **44**, 98–104
34. Ly, H. D., and Withers, S. G. (1999) Mutagenesis of glycosidases. *Annu. Rev. Biochem.* **68**, 487–522
35. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21
36. Vieille, C., and Zeikus, G. J. (2001) Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability. *Microbiol. Mol. Biol. Rev.* **65**, 1–43
37. Feller, G., and Gerday, C. (1997) Psychrophilic enzymes: molecular basis of cold adaptation. *Cell. Mol. Life Sci.* **53**, 830–841
38. Feller, G., and Gerday, C. (2003) Psychrophilic enzymes: hot topics in cold adaptation. *Nat. Rev. Microbiol.* **1**, 200–208
39. Aittaleb, M., Hubner, R., Lamotte-Brasseur, J., and Gerday, C. (1997) Cold adaptation parameters derived from cDNA sequencing and molecular modelling of elastase from Antarctic fish *Notothenia neglecta*. *Protein Eng.* **10**, 475–477
40. Russell, R. J., Gerike, U., Danson, M. J., Hough, D. W., and Taylor, G. L. (1998) Structural adaptations of the cold-active citrate synthase from an Antarctic bacterium. *Structure* **6**, 351–361
41. Jensen, R. A. (1976) Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol.* **30**, 409–425
42. Dorn, E., and Knackmuss, H. J. (1978) Chemical structure and biodegradability of halogenated aromatic compounds. Substituent effects on 1,2-dioxygenation of catechol. *Biochem. J.* **174**, 85–94

RÉFÉRENCES BIBLIOGRAPHIQUES

[A]

- Achong B. G., Mansell P. W., Epstein M. A. and Clifford P. (1971). "An unusual virus in cultures from a human nasopharyngeal carcinoma." J Natl Cancer Inst **46**(2): 299-307.
- Adams P. D., Afonine P. V., Bunkoczi G., Chen V. B., Davis I. W., Echols N., Headd J. J., Hung L.-W., Kapral G. J., Grosse-Kunstleve R. W., McCoy A. J., Moriarty N. W., Oeffner R., Read R. J., Richardson D. C., Richardson J. S., Terwilliger T. C. and Zwart P. H. (2010). "PHENIX: a comprehensive Python-based system for macromolecular structure solution doi:10.1107/S0907444909052925." Acta Crystallographica Section D **66**(2): 213-221.
- Adolph M. B., Webb J. and Chelico L. (2013). Retroviral Restriction Factor APOBEC3G Delays the Initiation of DNA Synthesis by HIV-1 Reverse Transcriptase. PLoS One. United States. **8**: e64196.
- Agapkina J., Smolov M., Barbe S., Zubin E., Zatsepin T., Deprez E., Le Bret M., Mouscadet J. F. and Gottikh M. (2006). "Probing of HIV-1 integrase/DNA interactions using novel analogs of viral DNA." J Biol Chem **281**(17): 11530-40.
- Aittaleb M., Hubner R., Lamotte-Brasseur J. and Gerday C. (1997). "Cold adaptation parameters derived from cDNA sequencing and molecular modelling of elastase from Antarctic fish *Notothenia neglecta*." Protein Eng **10**(5): 475-7.
- Al-Mawsawi L. Q., Fikkert V., Dayam R., Witvrouw M., Burke T. R., Jr., Borchers C. H. and Neamati N. (2006). Discovery of a small-molecule HIV-1 integrase inhibitor-binding site. Proc Natl Acad Sci U S A. United States. **103**: 10080-5.
- Amann E., Brosius J. and Ptashne M. (1983). "Vectors bearing a hybrid trp-lac promoter useful for regulated expression of cloned genes in *Escherichia coli*." Gene **25**(2-3): 167-78.
- Arts E. J. and Hazuda D. J. (2012). "HIV-1 antiretroviral drug therapy." Cold Spring Harb Perspect Med **2**(4): a007161.

[B]

- Balasubramaniam K. (1984). "Purification of alpha-galactosidase from coconut endosperm by affinity chromatography." Journal of the National Science Council of Sri Lanka **12**(1): 113-128.
- Ballandras A., Moreau K., Robert X., Confort M. P., Merceron R., Haser R., Ronfort C. and Gouet P. (2011). "A crystal structure of the catalytic core domain of an avian sarcoma and leukemia virus integrase suggests an alternate dimeric assembly." PLoS One **6**(8): e23032.
- Baltimore D. (1970). "RNA-dependent DNA polymerase in virions of RNA tumour viruses." Nature **226**(5252): 1209-11.
- Baltimore D. (1971). "Expression of animal virus genomes." Bacteriol Rev **35**(3): 235-41.
- Bannert N. and Kurth R. (2004). "Retroelements and the human genome: new perspectives on an old relation." Proc Natl Acad Sci U S A **101** Suppl 2: 14572-9.
- Bao K. K., Wang H., Miller J. K., Erie D. A., Skalka A. M. and Wong I. (2003). "Functional oligomeric state of avian sarcoma virus integrase." J Biol Chem **278**(2): 1323-7.
- Barre-Sinoussi F., Chermann J. C., Rey F., Nugeyre M. T., Chamaret S., Gruest J., Dauguet C., Axler-Blin C., Vezinet-Brun F., Rouzioux C., Rozenbaum W. and Montagnier L. (1983). "Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS)." Science **220**(4599): 868-71.
- Bernhard W. (1960). "The detection and study of tumor viruses with the electron microscope." Cancer Res **20**: 712-27.

- Bessette P. H., Aslund F., Beckwith J. and Georgiou G. (1999). "Efficient folding of proteins with multiple disulfide bonds in the *Escherichia coli* cytoplasm." Proc Natl Acad Sci U S A **96**(24): 13703-8.
- Bierry H. (1913). "Diastatic cleavage of glucosides and galactosides." Comptes rendus hebdomadaires de séances de la société de biologie: 156-265.
- Birnboim H. C. and Doly J. (1979). "A rapid alkaline extraction procedure for screening recombinant plasmid DNA." Nucleic Acids Res **7**(6): 1513-23.
- Bischerour J., Leh H., Deprez E., Brochon J. C. and Mouscadet J. F. (2003). "Disulfide-linked integrase oligomers involving C280 residues are formed in vitro and in vivo but are not essential for human immunodeficiency virus replication." J Virol **77**(1): 135-41.
- Bishop D. F. and Desnick R. J. (1981). "Affinity purification of alpha-galactosidase A from human spleen, placenta, and plasma with elimination of pyrogen contamination. Properties of the purified splenic enzyme compared to other forms." J Biol Chem **256**(3): 1307-16.
- Bittner J. J. (1936). "Some possible effects of nursing on the mammary gland tumour incidence in mice." Science **84**: 162-162.
- Blumenthal R., Durell S. and Viard M. (2012). "HIV entry and envelope glycoprotein-mediated fusion." J Biol Chem **287**(49): 40841-9.
- Bojja R. S., Andrade M. D., Merkel G., Weigand S., Dunbrack R. L., Jr. and Skalka A. M. (2013). "Architecture and Assembly of HIV Integrase Multimers in the Absence of DNA Substrates." J Biol Chem.
- Bojja R. S., Andrade M. D., Weigand S., Merkel G., Yarychivska O., Henderson A., Kummerling M. and Skalka A. M. (2011). "Architecture of a full-length retroviral integrase monomer and dimer, revealed by small angle X-ray scattering and chemical cross-linking." J Biol Chem **286**(19): 17047-59.
- Briones M. S., Dobard C. W. and Chow S. A. (2010). "Role of human immunodeficiency virus type 1 integrase in uncoating of the viral core." J Virol **84**(10): 5181-90.
- Brodin P., Pinskaya M., Buckle M., Parsch U., Romanova E., Engels J., Gottikh M. and Mouscadet J. F. (2002). "Disruption of HIV-1 integrase-DNA complexes by short 6-oxocytosine-containing oligonucleotides." Biochemistry **41**(5): 1529-38.
- Brodziak A., Ziolk E., Muc-Wierzgon M., Nowakowska-Zajdel E., Kokot T. and Klakla K. (2012). "The role of human endogenous retroviruses in the pathogenesis of autoimmune diseases." Med Sci Monit **18**(6): RA80-8.
- Brown J. R. and Doolittle W. F. (1997). "Archaea and the prokaryote-to-eukaryote transition." Microbiol Mol Biol Rev **61**(4): 456-502.
- Bruel L., Sulzenbacher G., Cervera Tison M., Pujol A., Nicoletti C., Perrier J., Galinier A., Ropartz D., Fons M., Pompeo F. and Giardina T. (2011). "alpha-Galactosidase/sucrose kinase (AgaSK), a novel bifunctional enzyme from the human microbiome coupling galactosidase and kinase activities." J Biol Chem **286**(47): 40814-23.
- Brugge J. S. and Erikson R. L. (1977). "Identification of a transformation-specific antigen induced by an avian sarcoma virus." Nature **269**(5626): 346-8.
- Brünger A. T. (1992). "Free R value: a novel statistical quantity for assessing the accuracy of crystal structures." NATURE **355**: 472-475.
- Bujacz G., Alexandratos J., Qing Z. L., Clement-Mella C. and Wlodawer A. (1996). "The catalytic domain of human immunodeficiency virus integrase: ordered active site in the F185H mutant." FEBS Lett **398**(2-3): 175-8.
- Bujacz G., Alexandratos J., Wlodawer A., Merkel G., Andrade M., Katz R. A. and Skalka A. M. (1997). "Binding of different divalent cations to the active site of avian sarcoma virus integrase and their effects on enzymatic activity." J Biol Chem **272**(29): 18161-8.
- Bujacz G., Jaskolski M., Alexandratos J., Wlodawer A., Merkel G., Katz R. A. and Skalka A. M. (1995). "High-resolution structure of the catalytic domain of avian sarcoma virus integrase." J Mol Biol **253**(2): 333-46.
- Bujacz G., Jaskolski M., Alexandratos J., Wlodawer A., Merkel G., Katz R. A. and Skalka A. M. (1996). "The catalytic domain of avian sarcoma virus integrase: conformation of the active-site residues in the presence of divalent cations." Structure **4**(1): 89-96.

Bushman F. D. and Wang B. (1994). "Rous sarcoma virus integrase protein: mapping functions for catalysis and substrate binding." J Virol **68**(4): 2215-23.

[C]

- Cai M., Zheng R., Caffrey M., Craigie R., Clore G. M. and Gronenborn A. M. (1997). "Solution structure of the N-terminal zinc binding domain of HIV-1 integrase." Nat Struct Biol **4**(7): 567-77.
- Cantarel B. L., Coutinho P. M., Rancurel C., Bernard T., Lombard V. and Henrissat B. (2009). "The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics." Nucleic Acids Res **37**(Database issue): D233-8.
- Carteau S., Gorelick R. J. and Bushman F. D. (1999). "Coupled integration of human immunodeficiency virus type 1 cDNA ends by purified integrase in vitro: stimulation by the viral nucleocapsid protein." J Virol **73**(8): 6670-9.
- Cavalluzzo C., Voet A., Christ F., Kumar Singh B., Sharma A., Debyser Z., De Maeyer M. and Van der Eycken E. (2012). "De novo design of small molecule inhibitors targeting the LEDGF/p75-HIV integrase interaction." RSC Advances **2**: 974-984.
- Chang C. F. (2012). Carbohydrates - Comprehensive Studies on Glycobiology and Glycotechnology, InTech.
- Chang J. Y. (1985). "Thrombin specificity. Requirement for apolar amino acids adjacent to the thrombin cleavage site of polypeptide substrate." Eur J Biochem **151**(2): 217-24.
- Chen J., Acton T. B., Basu S. K., Montelione G. T. and Inouye M. (2002). "Enhancement of the solubility of proteins overexpressed in *Escherichia coli* by heat shock." J Mol Microbiol Biotechnol **4**(6): 519-24.
- Chen J. C., Krucinski J., Miercke L. J., Finer-Moore J. S., Tang A. H., Leavitt A. D. and Stroud R. M. (2000). "Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding." Proc Natl Acad Sci U S A **97**(15): 8233-8.
- Cherepanov P. (2007). "LEDGF/p75 interacts with divergent lentiviral integrases and modulates their enzymatic activity in vitro." Nucleic Acids Res **35**(1): 113-24.
- Cherepanov P., Ambrosio A. L., Rahman S., Ellenberger T. and Engelman A. (2005). Structural basis for the recognition between HIV-1 integrase and transcriptional coactivator p75. Proc Natl Acad Sci U S A, United States. **102**: 17308-13.
- Cherepanov P., Maertens G., Proost P., Devreese B., Van Beeumen J., Engelborghs Y., De Clercq E. and Debyser Z. (2003). "HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells." J Biol Chem **278**(1): 372-81.
- Choi S. I., Song H. W., Moon J. W. and Seong B. L. (2001). "Recombinant enterokinase light chain with affinity tag: expression from *Saccharomyces cerevisiae* and its utilities in fusion protein technology." Biotechnol Bioeng **75**(6): 718-24.
- Chow S. A., Vincent K. A., Ellison V. and Brown P. O. (1992). "Reversal of integration and DNA splicing mediated by integrase of human immunodeficiency virus." Science **255**(5045): 723-6.
- Christ F., Voet A., Marchand A., Nicolet S., Desimmie B. A., Marchand D., Bardiot D., Van der Veken N. J., Van Remoortel B., Strelkov S. V., De Maeyer M., Chaltin P. and Debyser Z. (2010). Rational design of small-molecule inhibitors of the LEDGF/p75-integrase interaction and HIV replication. Nat Chem Biol, United States. **6**: 442-8.
- Clavel F., Guetard D., Brun-Vezinet F., Chamaret S., Rey M. A., Santos-Ferreira M. O., Laurent A. G., Dauguet C., Katlama C., Rouzioux C. and et al. (1986). "Isolation of a new human retrovirus from West African patients with AIDS." Science **233**(4761): 343-6.
- Coffin J., Haase A., Levy J. A., Montagnier L., Oroszlan S., Teich N., Temin H., Toyoshima K., Varmus H., Vogt P. and et al. (1986). "What to call the AIDS virus?" Nature **321**(6065): 10.
- Coffin J. M., Hughes S. H. and Varmus H. E. (1997). Retroviruses.
- Cohen J. (2002). Therapies. Confronting the limits of success. Science, United States. **296**: 2320-4.

- Coleman J., Eaton S., Merkel G., Skalka A. M. and Laue T. (1999). "Characterization of the self association of Avian sarcoma virus integrase by analytical ultracentrifugation." J Biol Chem **274**(46): 32842-6.
- Collaborative Computational Project N. (1994). "The CCP4 suite: programs for protein crystallography." Acta Crystallogr D Biol Crystallogr **50**(Pt 5): 760-3.
- Comfort D. A., Bobrov K. S., Ivanen D. R., Shabalin K. A., Harris J. M., Kulminskaya A. A., Brumer H. and Kelly R. M. (2007). "Biochemical analysis of *Thermotoga maritima* GH36 alpha-galactosidase (TmGalA) confirms the mechanistic commonality of clan GH-D glycoside hydrolases." Biochemistry **46**(11): 3319-30.
- Cox M. M., Doudna J. and O'Donnell M. (2011). Molecular Biology: Principles and Practice, W. H. Freeman.
- Cribier A., Segéral E., Delelis O., Parissi V., Simon A., Ruff M., Benarous R. and Emiliani S. (2011). Mutations affecting interaction of integrase with TNPO3 do not prevent HIV-1 cDNA nuclear import. Retrovirology. England, 2011 Cribier et al; licensee BioMed Central Ltd. **8**: 104.
- Crick F. (1970). "Central dogma of molecular biology." Nature **227**(5258): 561-3.
- Crick F. H. (1958). "On protein synthesis." Symp Soc Exp Biol **12**: 138-63.

[D]

- Dalgleish A. G., Beverley P. C., Clapham P. R., Crawford D. H., Greaves M. F. and Weiss R. A. (1984). "The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus." Nature **312**(5996): 763-7.
- Davanloo P., Rosenberg A. H., Dunn J. J. and Studier F. W. (1984). "Cloning and expression of the gene for bacteriophage T7 RNA polymerase." Proc Natl Acad Sci U S A **81**(7): 2035-9.
- Davies G. and Henrissat B. (1995). "Structures and mechanisms of glycosyl hydrolases." Structure **3**(9): 853-9.
- Davies G. J., Wilson K. S. and Henrissat B. (1997). "Nomenclature for sugar-binding subsites in glycosyl hydrolases." Biochem J **321** (Pt 2): 557-9.
- De Soultrait V. R., Caumont A., Parissi V., Morellet N., Ventura M., Lenoir C., Litvak S., Fournier M. and Roques B. (2002). "A novel short peptide is a specific inhibitor of the human immunodeficiency virus type 1 integrase." J Mol Biol **318**(1): 45-58.
- Dean K. J. and Sweeley C. C. (1979). "Studies on human liver alpha-galactosidases. III. Partial characterization of carbohydrate-binding specificities." J Biol Chem **254**(20): 10006-10.
- Demeulemeester J., Christ F., De Maeyer M. and Debyser Z. (2012). "Fueling HIV-1 integrase drug design with structural insights." Drug Discovery Today: Technologies **9**(3): 205-212.
- Demeulemeester J., Tintori C., Botta M., Debyser Z. and Christ F. (2012). Development of an AlphaScreen-based HIV-1 integrase dimerization assay for discovery of novel allosteric inhibitors. J Biomol Screen. United States. **17**: 618-28.
- Deprez E., Tauc P., Leh H., Mouscadet J. F., Auclair C. and Brochon J. C. (2000). "Oligomeric states of the HIV-1 integrase as measured by time-resolved fluorescence anisotropy." Biochemistry **39**(31): 9275-84.
- Dey P. M. and Pridham J. B. (1972). "Biochemistry of alpha-galactosidases." Adv Enzymol Relat Areas Mol Biol **36**: 91-130.
- Di Guan C., Li P., Riggs P. D. and Inouye H. (1988). "Vectors that facilitate the expression and purification of foreign peptides in *Escherichia coli* by fusion to maltose-binding protein." Gene **67**(1): 21-30.
- Di Stefano M., Miceli E., Gotti S., Missanelli A., Mazzocchi S. and Corazza G. R. (2007). "The effect of oral alpha-galactosidase on intestinal gas production and gas-related symptoms." Dig Dis Sci **52**(1): 78-83.
- Dolan J., Chen A., Weber I. T., Harrison R. W. and Leis J. (2009). "Defining the DNA substrate binding sites on HIV-1 integrase." J Mol Biol **385**(2): 568-79.

- Dorn E. and Knackmuss H. J. (1978). "Chemical structure and biodegradability of halogenated aromatic compounds. Substituent effects on 1,2-dioxygenation of catechol." Biochem J **174**(1): 85-94.
- Du L., Zhao Y. X., Yang L. M., Zheng Y. T., Tang Y., Shen X. and Jiang H. L. (2008). "Symmetrical 1-pyrrolidineacetamide showing anti-HIV activity through a new binding site on HIV-1 integrase." Acta Pharmacol Sin **29**(10): 1261-7.
- Duyk G., Leis J., Longiaru M. and Skalka A. M. (1983). "Selective cleavage in the avian retroviral long terminal repeat sequence by the endonuclease associated with the alpha beta form of avian reverse transcriptase." Proc Natl Acad Sci U S A **80**(22): 6745-9.
- Dyda F., Hickman A. B., Jenkins T. M., Engelman A., Craigie R. and Davies D. R. (1994). "Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases." Science **266**(5193): 1981-6.

[E]

- Eijkelenboom A. P., Lutzke R. A., Boelens R., Plasterk R. H., Kaptein R. and Hard K. (1995). "The DNA-binding domain of HIV-1 integrase has an SH3-like fold." Nat Struct Biol **2**(9): 807-10.
- Eijkelenboom A. P., Sprangers R., Hard K., Puras Lutzke R. A., Plasterk R. H., Boelens R. and Kaptein R. (1999). "Refined solution structure of the C-terminal DNA-binding domain of human immunovirus-1 integrase." Proteins **36**(4): 556-64.
- Eijkelenboom A. P., van den Ent F. M., Vos A., Doreleijers J. F., Hard K., Tullius T. D., Plasterk R. H., Kaptein R. and Boelens R. (1997). "The solution structure of the amino-terminal HHCC domain of HIV-2 integrase: a three-helix bundle stabilized by zinc." Curr Biol **7**(10): 739-46.
- Ellermann V. and Bang O. (1908). "Experimentelle Leukämie bei Hühnern." Zentralbl. Bakteriol. Parasitenkd. Infektionskr. Hyg. Abt. Orig. **46**: 595-609.
- Emiliani S., Mousnier A., Busschots K., Maroun M., Van Maele B., Tempe D., Vandekerckhove L., Moisan F., Ben-Slama L., Witvrouw M., Christ F., Rain J. C., Dargemont C., Debyser Z. and Benarous R. (2005). "Integrase mutants defective for interaction with LEDGF/p75 are impaired in chromosome tethering and HIV-1 replication." J Biol Chem **280**(27): 25517-23.
- Emsley P. and Cowtan K. (2004). "Coot: model-building tools for molecular graphics." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2126-32.
- Eneyskaya E. V., Golubev A. M., Kachurin A. M., Savel'ev A. N. and Neustroev K. N. (1997). "Transglycosylation activity of alpha-D-galactosidase from *Trichoderma reesei*. An investigation of the active site." Carbohydr Res **305**(1): 83-91.
- Engelman A. (1999). "In vivo analysis of retroviral integrase structure and function." Adv Virus Res **52**: 411-26.
- Engelman A. and Cherepanov P. (2008). "The lentiviral integrase binding protein LEDGF/p75 and HIV-1 replication." PLoS Pathog **4**(3): e1000046.
- Esposito D. and Craigie R. (1998). "Sequence specificity of viral end DNA binding by HIV-1 integrase reveals critical regions for protein-DNA interaction." EMBO J **17**(19): 5832-43.

[F]

- Farnet C. M. and Haseltine W. A. (1990). "Integration of human immunodeficiency virus type 1 DNA in vitro." Proc Natl Acad Sci U S A **87**(11): 4164-8.
- Faschinger A., Rouault F., Sollner J., Lukas A., Salmons B., Gunzburg W. H. and Indik S. (2008). "Mouse mammary tumor virus integration site selection in human and mouse genomes. J Virol. United States. **82**: 1360-7.

- Faure A., Calmels C., Desjobert C., Castroviejo M., Caumont-Sarcos A., Tarrago-Litvak L., Litvak S. and Parissi V. (2005). "HIV-1 integrase crosslinked oligomers are active in vitro." Nucleic Acids Res **33**(3): 977-86.
- Feldt-Rasmussen U., Rasmussen A. K., Mersebach H., Rosenberg K. M., Hasholt L. and Sorensen S. A. (2002). "Fabry disease--a metabolic disorder with a challenge for endocrinologists?" Horm Res **58**(6): 259-65.
- Feller G. and Gerday C. (2003). "Psychrophilic enzymes: hot topics in cold adaptation." Nat Rev Microbiol **1**(3): 200-8.
- Feng Y., Broder C. C., Kennedy P. E. and Berger E. A. (1996). "HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor." Science **272**(5263): 872-7.
- Fernandez-Leiro R., Pereira-Rodriguez A., Cerdan M. E., Becerra M. and Sanz-Aparicio J. (2010). "Structural analysis of *Saccharomyces cerevisiae* alpha-galactosidase and its complexes with natural substrates reveals new insights into substrate specificity of GH27 glycosidases." J Biol Chem **285**(36): 28020-33.
- Fitzgerald P. M., Wu J. K. and Toney J. H. (1998). Unanticipated inhibition of the metallo-beta-lactamase from *Bacteroides fragilis* by 4-morpholineethanesulfonic acid (MES): a crystallographic study at 1.85-A resolution. Biochemistry. United States. **37**: 6791-800.
- Fitzkee N. C., Masse J. E., Shen Y., Davies D. R. and Bax A. (2010). "Solution conformation and dynamics of the HIV-1 integrase core domain." J Biol Chem **285**(23): 18072-84.
- Foucault M., Watzlawick H., Mattes R., Haser R. and Gouet P. (2006). "Crystallization and preliminary X-ray diffraction studies of two thermostable alpha-galactosidases from glycoside hydrolase family 36." Acta Crystallogr Sect F Struct Biol Cryst Commun **62**(Pt 2): 100-3.
- Franke D. and Svergun D. I. (2009). "DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering." Journal of Applied Crystallography **42**: 342-346.
- Fraser-Reid B. O., Tatsuta K. and Thiem J. (2001). Glycoscience: Chemistry & Chemical Biology, Springer.
- Fredslund F., Hachem M. A., Larsen R. J., Sorensen P. G., Coutinho P. M., Lo Leggio L. and Svensson B. (2011). "Crystal structure of alpha-galactosidase from *Lactobacillus acidophilus* NCFM: insight into tetramer formation and substrate binding." J Mol Biol **412**(3): 466-80.
- Fridjonsson O., Watzlawick H., Gehweiler A. and Mattes R. (1999). "Thermostable alpha-galactosidase from *Bacillus stearothermophilus* NUB3621: cloning, sequencing and characterization." FEMS Microbiol Lett **176**(1): 147-53.
- Fridjonsson O., Watzlawick H. and Mattes R. (2002). "Thermoadaptation of alpha-galactosidase AgaB1 in *Thermus thermophilus*." J Bacteriol **184**(12): 3385-91.
- Fujimoto Z., Kaneko S., Momma M., Kobayashi H. and Mizuno H. (2003). "Crystal structure of rice alpha-galactosidase complexed with D-galactose." J Biol Chem **278**(22): 20313-8.

[G]

- Gallo R. C., Salahuddin S. Z., Popovic M., Shearer G. M., Kaplan M., Haynes B. F., Palker T. J., Redfield R., Oleske J., Safai B. and et al. (1984). "Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS." Science **224**(4648): 500-3.
- Ganter C., Böck A., Buckel P. and Mattes R. (1988). "Production of thermostable, recombinant alpha-galactosidase suitable for raffinose elimination from sugar beet syrup." Journal of Biotechnology **8**: 301-310.
- Gao K., Gorelick R. J., Johnson D. G. and Bushman F. (2003). "Cofactors for human immunodeficiency virus type 1 cDNA integration in vitro." J Virol **77**(2): 1598-603.

- Garman S. C. and Garboczi D. N. (2004). "The molecular defect leading to Fabry disease: structure of human alpha-galactosidase." *J Mol Biol* **337**(2): 319-35.
- Garman S. C., Hannick L., Zhu A. and Garboczi D. N. (2002). "The 1.9 Å structure of alpha-N-acetylgalactosaminidase: molecular basis of glycosidase deficiency diseases." *Structure* **10**(3): 425-34.
- Geier G. E. and Modrich P. (1979). "Recognition sequence of the dam methylase of *Escherichia coli* K12 and mode of cleavage of Dpn I endonuclease." *J Biol Chem* **254**(4): 1408-13.
- Gerday C., Aittaleb M., Bentahir M., Chessa J. P., Claverie P., Collins T., D'Amico S., Dumont J., Garsoux G., Georgette D., Hoyoux A., Lonhienne T., Meuwis M. A. and Feller G. (2000). "Cold-adapted enzymes: from fundamentals to biotechnology." *Trends Biotechnol* **18**(3): 103-7.
- Goldgur Y., Craigie R., Cohen G. H., Fujiwara T., Yoshinaga T., Fujishita T., Sugimoto H., Endo T., Murai H. and Davies D. R. (1999). "Structure of the HIV-1 integrase catalytic domain complexed with an inhibitor: a platform for antiviral drug design." *Proc Natl Acad Sci U S A* **96**(23): 13040-3.
- Goldgur Y., Dyda F., Hickman A. B., Jenkins T. M., Craigie R. and Davies D. R. (1998). "Three new structures of the core domain of HIV-1 integrase: an active site that binds magnesium." *Proc Natl Acad Sci U S A* **95**(16): 9150-4.
- Goldstein J. (1989). "Conversion of ABO blood groups." *Transfus Med Rev* **3**(3): 206-12.
- Golubev A. M., Nagem R. A., Brandao Neto J. R., Neustroev K. N., Eneyskaya E. V., Kulminskaya A. A., Shabalin K. A., Savel'ev A. N. and Polikarpov I. (2004). "Crystal structure of alpha-galactosidase from *Trichoderma reesei* and its complex with galactose: implications for catalytic mechanism." *J Mol Biol* **339**(2): 413-22.
- Gomes J. and Steiner W. (2004). "The Biocatalytic Potential of Extremophiles and Extremozymes." *Food Technology and Biotechnology* **42**(4): 223-235.
- Gramberg T., Sunseri N. and Landau N. R. (2009). "Accessories to the crime: recent advances in HIV accessory protein biology." *Curr HIV/AIDS Rep* **6**(1): 36-42.
- Grandgenett D. P., Vora A. C. and Schiff R. D. (1978). "A 32,000-dalton nucleic acid-binding protein from avian reovirus cores possesses DNA endonuclease activity." *Virology* **89**(1): 119-32.
- Grassmann R., Aboud M. and Jeang K. T. (2005). Molecular mechanisms of cellular transformation by HTLV-1 Tax. *Oncogene*, England. **24**: 5976-85.
- Greenwald J., Le V., Butler S. L., Bushman F. D. and Choe S. (1999). "The mobility of an HIV-1 integrase active site loop is correlated with catalytic activity." *Biochemistry* **38**(28): 8892-8.
- Groom H. C., Anderson E. C. and Lever A. M. (2009). "Rev: beyond nuclear export." *J Gen Virol* **90**(Pt 6): 1303-18.
- Gross L. (1957). "Development and serial cellfree passage of a highly potent strain of mouse leukemia virus." *Proc Soc Exp Biol Med* **94**(4): 767-71.
- Guce A. I., Clark N. E., Salgado E. N., Ivanen D. R., Kulminskaya A. A., Brumer H., 3rd and Garman S. C. (2010). "Catalytic mechanism of human alpha-galactosidase." *J Biol Chem* **285**(6): 3625-32.
- Gupta K., Curtis J. E., Krueger S., Hwang Y., Cherepanov P., Bushman F. D. and Van Dyne G. D. (2012). "Solution conformations of prototype foamy virus integrase and its stable synaptic complex with U5 viral DNA." *Structure* **20**(11): 1918-28.

[H]

- Hare S., Di Nunzio F., Labeja A., Wang J., Engelman A. and Cherepanov P. (2009). "Structural basis for functional tetramerization of lentiviral integrase." *PLoS Pathog* **5**(7): e1000515.
- Hare S., Gupta S. S., Valkov E., Engelman A. and Cherepanov P. (2010). "Retroviral intasome assembly and inhibition of DNA strand transfer." *Nature* **464**(7286): 232-6.

- Hare S., Maertens G. N. and Cherepanov P. (2012). 3'-processing and strand transfer catalysed by retroviral integrase in crystallo. *EMBO J.* England. **31**: 3020-8.
- Hare S., Shun M. C., Gupta S. S., Valkov E., Engelman A. and Cherepanov P. (2009). "A novel co-crystal structure affords the design of gain-of-function lentiviral integrase mutants in the presence of modified PSIP1/LEDGF/p75." *PLoS Pathog* **5**(1): e1000259.
- Hartley J. L., Temple G. F. and Brasch M. A. (2000). "DNA cloning using in vitro site-specific recombination." *Genome Res* **10**(11): 1788-95.
- Hayouka Z., Rosenbluh J., Levin A., Loya S., Lebendiker M., Veprintsev D., Kotler M., Hizi A., Loyter A. and Friedler A. (2007). "Inhibiting HIV-1 integrase by shifting its oligomerization equilibrium." *Proc Natl Acad Sci U S A* **104**(20): 8316-21.
- Hayward W. S., Neel B. G. and Astrin S. M. (1981). "Activation of a cellular onc gene by promoter insertion in ALV-induced lymphoid leukemia." *Nature* **290**(5806): 475-80.
- Hazuda D. J. (2012). "HIV integrase as a target for antiretroviral therapy." *Curr Opin HIV AIDS* **7**(5): 383-9.
- Hefferon K. (2006). *Recent advances in RNA Virus Replication*.
- Henrissat B. (1991). "A classification of glycosyl hydrolases based on amino acid sequence similarities." *Biochem J* **280** (Pt 2): 309-16.
- Henrissat B. and Bairoch A. (1993). "New families in the classification of glycosyl hydrolases based on amino acid sequence similarities." *Biochem J* **293** (Pt 3): 781-8.
- Henrissat B. and Bairoch A. (1996). "Updating the sequence-based classification of glycosyl hydrolases." *Biochem J* **316** (Pt 2): 695-6.
- Heuer T. S. and Brown P. O. (1997). "Mapping features of HIV-1 integrase near selected sites on viral and target DNA molecules in an active enzyme-DNA complex by photo-cross-linking." *Biochemistry* **36**(35): 10655-65.
- Heuer T. S. and Brown P. O. (1998). "Photo-cross-linking studies suggest a model for the architecture of an active human immunodeficiency virus type 1 integrase-DNA complex." *Biochemistry* **37**(19): 6667-78.
- Hidaka M., Kitaoka M., Hayashi K., Wakagi T., Shoun H. and Fushinobu S. (2006). "Structural dissection of the reaction mechanism of cellobiose phosphorylase." *Biochem J* **398**(1): 37-43.
- Hippenmeyer P. J. and Grandgenett D. P. (1984). "Requirement of the avian retrovirus pp32 DNA binding protein domain for replication." *Virology* **137**(2): 358-70.
- Hong W. K., Bast R. C., Hait W., Kufe D. W., Pollock R. E., Weichselbaum R. R., Holland J. F. and Iii E. F. (2010). *Holland-Frei Cancer Medicine* **8**, McGraw-Hill Medical.
- Hough D. W. and Danson M. J. (1999). "Extremozymes." *Curr Opin Chem Biol* **3**(1): 39-46.

[I]

- Ikeda T., Nishitsuji H., Zhou X., Nara N., Ohashi T., Kannagi M. and Masuda T. (2004). "Evaluation of the functional involvement of human immunodeficiency virus type 1 integrase in nuclear import of viral cDNA during acute infection." *J Virol* **78**(21): 11563-73.

[J]

- Jacque J. M. and Stevenson M. (2006). "The inner-nuclear-envelope protein emerlin regulates HIV-1 infectivity." *Nature* **441**(7093): 641-5.
- Jaskolski M., Alexandratos J. N., Bujacz G. and Wlodawer A. (2009). "Piecing together the structure of retroviral integrase, an important target in AIDS therapy." *FEBS J* **276**(11): 2926-46.
- Jensen R. A. (1976). "Enzyme recruitment in evolution of new function." *Annu Rev Microbiol* **30**: 409-25.
- Jilaveanu L. B. and Oliver D. (2006). SecA dimer cross-linked at its subunit interface is functional for protein translocation. *J Bacteriol.* United States. **188**: 335-8.

Jones K. S., Coleman J., Merkel G. W., Laue T. M. and Skalka A. M. (1992). "Retroviral integrase functions as a multimer and can turn over catalytically." J Biol Chem **267**(23): 16037-40.

[K]

Kabsch W. (2010). Integration, scaling, space-group assignment and post-refinement. Acta Crystallogr D Biol Crystallogr. England. **66**: 133-44.

Kalpana G. V., Reicin A., Cheng G. S., Sorin M., Paik S. and Goff S. P. (1999). "Isolation and characterization of an oligomerization-negative mutant of HIV-1 integrase." Virology **259**(2): 274-85.

Kalyanaraman V. S., Sarngadharan M. G., Robert-Guroff M., Miyoshi I., Golde D. and Gallo R. C. (1982). "A new subtype of human T-cell leukemia virus (HTLV-II) associated with a T-cell variant of hairy cell leukemia." Science **218**(4572): 571-3.

Kapust R. B., Tozser J., Fox J. D., Anderson D. E., Cherry S., Copeland T. D. and Waugh D. S. (2001). "Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency." Protein Eng **14**(12): 993-1000.

Karn J. and Stoltzfus C. M. (2012). Transcriptional and posttranscriptional regulation of HIV-1 gene expression. Cold Spring Harb Perspect Med. United States. **2**: a006916.

Katz R. A., Merkel G., Kulkosky J., Leis J. and Skalka A. M. (1990). "The avian retroviral IN protein is both necessary and sufficient for integrative recombination in vitro." Cell **63**(1): 87-95.

Katz R. A. and Skalka A. M. (1988). "A C-terminal domain in the avian sarcoma-leukosis virus pol gene product is not essential for viral replication." J Virol **62**(2): 528-33.

Katzman M. and Sudol M. (1996). "Nonspecific alcoholysis, a novel endonuclease activity of human immunodeficiency virus type 1 and other retroviral integrases." J Virol **70**(4): 2598-604.

Kessl J. J., Eidahl J. O., Shkriabai N., Zhao Z., McKee C. J., Hess S., Burke T. R., Jr. and Kvaratskhelia M. (2009). An allosteric mechanism for inhibiting HIV-1 integrase with a small molecule. Mol Pharmacol. United States. **76**: 824-32.

Kessl J. J., Jena N., Koh Y., Taskent-Sezgin H., Slaughter A., Feng L., de Silva S., Wu L., Le Grice S. F., Engelman A., Fuchs J. R. and Kvaratskhelia M. (2012). "Multimode, cooperative mechanism of action of allosteric HIV-1 integrase inhibitors." J Biol Chem **287**(20): 16801-11.

King A. M. Q., Lefkowitz E., Adams M. J. and Carstens E. B. (2011). Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses, Elsevier.

Knochel T. R., Hennig M., Merz A., Darimont B., Kirschner K. and Jansonius J. N. (1996). The crystal structure of indole-3-glycerol phosphate synthase from the hyperthermophilic archaeon *Sulfolobus solfataricus* in three different crystal forms: effects of ionic strength. J Mol Biol. England. **262**: 502-15.

Koshland D. E. (1953). "STEREOCHEMISTRY AND THE MECHANISM OF ENZYMATIC REACTIONS." Biological Reviews **28**(4): 416-436.

Krishnan L. and Engelman A. (2012). "Retroviral integrase proteins and HIV-1 DNA integration." J Biol Chem **287**(49): 40858-66.

Kruskall M. S., AuBuchon J. P., Anthony K. Y., Herschel L., Pickard C., Biehl R., Horowitz M., Brambilla D. J. and Popovsky M. A. (2000). "Transfusion to blood group A and O patients of group B RBCs that have been enzymatically converted to group O." Transfusion **40**(11): 1290-8.

Kulkosky J., Katz R. A., Merkel G. and Skalka A. M. (1995). "Activities and substrate specificity of the evolutionarily conserved central domain of retroviral integrase." Virology **206**(1): 448-56.

Kurth R. and Bannert N. (2010). Retroviruses: Molecular Biology, Genomics and Pathogenesis, Caister Academic Press.

[L]

- Laemmli U. K. (1970). "Cleavage of structural proteins during the assembly of the head of bacteriophage T4." *Nature* **227**(5259): 680-5.
- Laguette N., Sobhian B., Casartelli N., Ringard M., Chable-Bessia C., Segal E., Yatim A., Emiliani S., Schwartz O. and Benkirane M. (2011). SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. *Nature*. England. **474**: 654-7.
- Le Grice S. F. (2012). "Human immunodeficiency virus reverse transcriptase: 25 years of research, drug discovery, and promise." *J Biol Chem* **287**(49): 40850-7.
- Lee S. K., Potempa M. and Swanstrom R. (2012). "The choreography of HIV-1 proteolytic processing and virion assembly." *J Biol Chem* **287**(49): 40867-74.
- Lee S. P., Xiao J., Knutson J. R., Lewis M. S. and Han M. K. (1997). "Zn²⁺ promotes the self-association of human immunodeficiency virus type-1 integrase in vitro." *Biochemistry* **36**(1): 173-80.
- Lee S. Y., Letts J. A. and Mackinnon R. (2008). Dimeric subunit stoichiometry of the human voltage-dependent proton channel Hv1. *Proc Natl Acad Sci U S A*. United States. **105**: 7692-5.
- Leis J., Baltimore D., Bishop J. M., Coffin J., Fleissner E., Goff S. P., Oroszlan S., Robinson H., Skalka A. M., Temin H. M. and et al. (1988). "Standardized and simplified nomenclature for proteins common to all retroviruses." *J Virol* **62**(5): 1808-9.
- Lenny L. L., Hurst R., Goldstein J. and Galbraith R. A. (1994). "Transfusions to group O subjects of 2 units of red cells enzymatically converted from group B to group O." *Transfusion* **34**(3): 209-14.
- Lesbats P., Botbol Y., Chevereau G., Vaillant C., Calmels C., Arneodo A., Andreola M. L., Lavigne M. and Parissi V. (2011). "Functional coupling between HIV-1 integrase and the SWI/SNF chromatin remodeling complex for efficient in vitro integration into stable nucleosomes." *PLoS Pathog* **7**(2): e1001280.
- Lesley S. A., Kuhn P., Godzik A., Deacon A. M., Mathews I., Kreusch A., Spraggon G., Klock H. E., McMullan D., Shin T., Vincent J., Robb A., Brinen L. S., Miller M. D., McPhillips T. M., Miller M. A., Scheibe D., Canaves J. M., Guda C., Jaroszewski L., Selby T. L., Elsliger M. A., Wooley J., Taylor S. S., Hodgson K. O., Wilson I. A., Schultz P. G. and Stevens R. C. (2002). "Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline." *Proc Natl Acad Sci U S A* **99**(18): 11664-9.
- Letvin N. L., Daniel M. D., Sehgal P. K., Desrosiers R. C., Hunt R. D., Waldron L. M., MacKey J. J., Schmidt D. K., Chalifoux L. V. and King N. W. (1985). "Induction of AIDS-like disease in macaque monkeys with T-cell tropic retrovirus STLV-III." *Science* **230**(4721): 71-3.
- Lewinski M. K., Yamashita M., Emerman M., Ciuffi A., Marshall H., Crawford G., Collins F., Shinn P., Leipzig J., Hannenhalli S., Berry C. C., Ecker J. R. and Bushman F. D. (2006). "Retroviral DNA integration: viral and cellular determinants of target-site selection." *PLoS Pathog* **2**(6): e60.
- Li H. Y., Zawahir Z., Song L. D., Long Y. Q. and Neamati N. (2006). "Sequence-based design and discovery of peptide inhibitors of HIV-1 integrase: insight into the binding mode of the enzyme." *J Med Chem* **49**(15): 4477-86.
- Li L., Li H. S., Pauza C. D., Bukrinsky M. and Zhao R. Y. (2005). "Roles of HIV-1 auxiliary proteins in viral pathogenesis and host-pathogen interactions." *Cell Res* **15**(11-12): 923-34.
- Lipford J. R., Worland S. T. and Farnet C. M. (1994). "Nucleotide binding by the HIV-1 integrase protein in vitro." *J Acquir Immune Defic Syndr* **7**(12): 1215-23.
- Liu Q. P., Sulzenbacher G., Yuan H., Bennett E. P., Pietz G., Saunders K., Spence J., Nudelman E., Lavery S. B., White T., Neveu J. M., Lane W. S., Bourne Y., Olsson M. L., Henrissat B. and Clausen H. (2007). "Bacterial glycosidases for the production of universal red blood cells." *Nat Biotechnol* **25**(4): 454-64.

- Lo M. C., Aulabaugh A., Jin G., Cowling R., Bard J., Malamas M. and Ellestad G. (2004). "Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery." Anal Biochem **332**(1): 153-9.
- Lodi P. J., Ernst J. A., Kuszewski J., Hickman A. B., Engelman A., Craigie R., Clore G. M. and Gronenborn A. M. (1995). "Solution structure of the DNA binding domain of HIV-1 integrase." Biochemistry **34**(31): 9826-33.
- Long D. and Yang D. (2009). Buffer interference with protein dynamics: a case study on human liver fatty acid binding protein. Biophys J. United States. **96**: 1482-8.
- Lubkowski J., Dauter Z., Yang F., Alexandratos J., Merkel G., Skalka A. M. and Wlodawer A. (1999). "Atomic Resolution Structures of the Core Domain of Avian Sarcoma Virus Integrase and Its D64N Mutant." Biochemistry **38**(45): 15060.
- Lubkowski J., Yang F., Alexandratos J., Merkel G., Katz R. A., Gravuer K., Skalka A. M. and Wlodawer A. (1998). "Structural basis for inactivating mutations and pH-dependent activity of avian sarcoma virus integrase." J Biol Chem **273**(49): 32685-9.
- Lubkowski J., Yang F., Alexandratos J., Wlodawer A., Zhao H., Burke T. R., Jr., Neamati N., Pommier Y., Merkel G. and Skalka A. M. (1998). "Structure of the catalytic domain of avian sarcoma virus integrase with a bound HIV-1 integrase-targeted inhibitor." Proc Natl Acad Sci U S A **95**(9): 4831-6.
- Lutzke R. A. and Plasterk R. H. (1998). "Structure-based mutational analysis of the C-terminal DNA-binding domain of human immunodeficiency virus type 1 integrase: critical residues for protein oligomerization and DNA binding." J Virol **72**(6): 4841-8.

[M]

- Maeda N., Fan H. and Yoshikai Y. (2008). "Oncogenesis by retroviruses: old and new paradigms." Rev Med Virol **18**(6): 387-405.
- Maertens G., Cherepanov P., Pluymers W., Busschots K., De Clercq E., Debyser Z. and Engelborghs Y. (2003). "LEDGF/p75 is essential for nuclear and chromosomal targeting of HIV-1 integrase in human cells." J Biol Chem **278**(35): 33528-39.
- Maertens G. N., Hare S. and Cherepanov P. (2010). "The mechanism of retroviral integration from X-ray structures of its key intermediates." Nature **468**(7321): 326-9.
- Maignan S., Guilloteau J. P., Zhou-Liu Q., Clement-Mella C. and Mikol V. (1998). "Crystal structures of the catalytic domain of HIV-1 integrase free and complexed with its metal cofactor: high level of similarity of the active site with other viral integrases." J Mol Biol **282**(2): 359-68.
- Maillot B., Levy N., Eiler S., Crucifix C., Granger F., Richert L., Didier P., Godet J., Pradeau-Aubretton K., Emiliani S., Nazabal A., Lesbats P., Parissi V., Mely Y., Moras D., Schultz P. and Ruff M. (2013). Structural and functional role of IN1 and LEDGF in the HIV-1 preintegration complex. PLoS One. United States. **8**: e60734.
- Malet I., Calvez V. and Marcelin A. G. (2012). "The future of integrase inhibitors of HIV-1." Curr Opin Virol **2**(5): 580-7.
- Malim M. H. and Bieniasz P. D. (2012). HIV Restriction Factors and Mechanisms of Evasion. Cold Spring Harb Perspect Med. United States. **2**: a006940.
- Malim M. H. and Emerman M. (2008). "HIV-1 accessory proteins--ensuring viral survival in a hostile environment." Cell Host Microbe **3**(6): 388-98.
- Maréchal V. and Piolot T. (1999). "Oncogenèse et cycle biologique des virus." Virologie **3**(4): 297-308.
- Markovic S. and Dutzler R. (2007). The structure of the cytoplasmic domain of the chloride channel ClC-Ka reveals a conserved interaction interface. Structure. United States. **15**: 715-25.
- Maroun R. G., Gayet S., Benleulmi M. S., Porumb H., Zargarian L., Merad H., Leh H., Mouscadet J. F., Troalen F. and Fermandjian S. (2001). "Peptide inhibitors of HIV-1 integrase dissociate the enzyme oligomers." Biochemistry **40**(46): 13840-8.

- Maroun R. G., Krebs D., Roshani M., Porumb H., Auclair C., Troalen F. and Femandjian S. (1999). "Conformational aspects of HIV-1 integrase inhibition by a peptide derived from the enzyme central domain and by antibodies raised against this peptide." Eur J Biochem **260**(1): 145-55.
- Marx J. L. (1985). "A virus by any other name." Science **227**(4693): 1449-51.
- Masuda T. (2011). "Non-Enzymatic Functions of Retroviral Integrase: The Next Target for Novel Anti-HIV Drug Development." Front Microbiol **2**: 210.
- Mazumder A., Neamati N., Sommadossi J. P., Gosselin G., Schinazi R. F., Imbach J. L. and Pommier Y. (1996). "Effects of nucleotide analogues on human immunodeficiency virus type 1 integrase." Mol Pharmacol **49**(4): 621-8.
- McCord M., Stahl S. J., Mueser T. C., Hyde C. C., Vora A. C., Grandgenett D. P. (1998). Purification of recombinant Rous Sarcoma Virus Integrase pocessing physical and catalytic properties similar to virion-derived integrase. Protein Expr Purif **14**(2):167-77.
- McLellan T. (1982). "Electrophoresis buffers for polyacrylamide gels at various pH." Anal Biochem **126**(1): 94-9.
- Merceron R., Foucault M., Haser R., Mattes R., Watzlawick H. and Gouet P. (2012). "The molecular mechanism of thermostable alpha-galactosidases AgaA and AgaB explained by x-ray crystallography and mutational studies." J Biol Chem **287**(47): 39642-52.
- Michel F., Crucifix C., Granger F., Eiler S., Mouscadet J. F., Korolev S., Agapkina J., Ziganshin R., Gottikh M., Nazabal A., Emiliani S., Benarous R., Moras D., Schultz P. and Ruff M. (2009). "Structural basis for HIV-1 DNA integration in the human genome, role of the LEDGF/P75 cofactor." EMBO J **28**(7): 980-91.
- Miroux B. and Walker J. E. (1996). Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. J Mol Biol. England. **260**: 289-98.
- Moalic Y., Blanchard Y., Felix H. and Jestin A. (2006). "Porcine endogenous retrovirus integration sites in the human genome: features in common with those of murine leukemia virus." J Virol **80**(22): 10980-8.
- Mohammed K. D., Topper M. B. and Muesing M. A. (2011). "Sequential deletion of the integrase (Gag-Pol) carboxyl terminus reveals distinct phenotypic classes of defective HIV-1." J Virol **85**(10): 4654-66.
- Molteni V., Greenwald J., Rhodes D., Hwang Y., Kwiatkowski W., Bushman F. D., Siegel J. S. and Choe S. (2001). "Identification of a small-molecule binding site at the dimer interface of the HIV integrase catalytic domain." Acta Crystallogr D Biol Crystallogr **57**(Pt 4): 536-44.
- Monis B. and Wasserkrug H. (1967). "Histochemistry of glycosidases of megakaryocytes and platelets. A comparative study." Histochemie **10**(4): 362-8.
- Moreau K., Faure C., Violot S., Verdier G. and Ronfort C. (2003). "Mutations in the C-terminal domain of ALSV (Avian Leukemia and Sarcoma Viruses) integrase alter the concerted DNA integration process in vitro." Eur J Biochem **270**(22): 4426-38.
- Mouscadet J. F., Delelis O., Marcelin A. G. and Tchertanov L. (2010). Resistance to HIV-1 integrase inhibitors: A structural perspective. Drug Resist Updat. Scotland, 2010 Elsevier Ltd. **13**: 139-50.
- Mulder L. C., Chakrabarti L. A. and Muesing M. A. (2002). Interaction of HIV-1 integrase with DNA repair protein hRad18. J Biol Chem. United States. **277**: 27489-93.
- Mullis K., Faloona F., Scharf S., Saiki R., Horn G. and Erlich H. (1986). "Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction." Cold Spring Harb Symp Quant Biol **51 Pt 1**: 263-73.
- Murphy R. A. and Power R. F. (2002). "Expression of an alpha-galactosidase from *Saccharomyces cerevisiae* in *Aspergillus awamori* and *Aspergillus oryzae*." J Ind Microbiol Biotechnol **28**(2): 97-102.

[N]

- Nakai H., Baumann M. J., Petersen B. O., Westphal Y., Hachem M. A., Dilokpimol A., Duus J. O., Schols H. A. and Svensson B. (2010). "*Aspergillus nidulans* alpha-galactosidase of glycoside hydrolase family 36 catalyses the formation of alpha-galacto-oligosaccharides by transglycosylation." FEBS J **277**(17): 3538-51.
- Narezkina A., Taganov K. D., Litwin S., Stoyanova R., Hayashi J., Seeger C., Skalka A. M. and Katz R. A. (2004). "Genome-wide analyses of avian sarcoma virus integration sites." J Virol **78**(21): 11656-63.
- Neamati N. and Wang B. (2011). HIV-1 Integrase: Mechanism and Inhibitor Design, Wiley.
- Neil S. J., Zang T. and Bieniasz P. D. (2008). Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. Nature, England. **451**: 425-30.
- Novy R., Drott D., Yaeger K. and Mierendorf R. (2001). "Overcoming codon bias of *E. coli* for enhanced protein expression." inNovations **12**: 1-3.
- Nunnari G., Smith J. A. and Daniel R. (2008). HIV-1 Tat and AIDS-associated cancer: targeting the cellular anti-cancer barrier? J Exp Clin Cancer Res, Italy. **27**: 3.

[O]

- Oren A. (2010). "Industrial and environmental applications of halophilic microorganisms." Environ Technol **31**(8-9): 825-34.
- Ornstein L. (1964). "Disc electrophoresis. I. Background and theory." Ann N Y Acad Sci **121**: 321-49.
- Oz I., Avidan O. and Hizi A. (2002). "Inhibition of the integrases of human immunodeficiency viruses type 1 and type 2 by reverse transcriptases." Biochem J **361**(Pt 3): 557-66.

[P]

- Pace C. N., Vajdos F., Fee L., Grimsley G. and Gray T. (1995). "How to measure and predict the molar absorption coefficient of a protein." Protein Sci **4**(11): 2411-23.
- Panganiban A. T. and Temin H. M. (1983). "The terminal nucleotides of retrovirus DNA are required for integration but not virus production." Nature **306**(5939): 155-60.
- Pantoliano M. W., Petrella E. C., Kwasnoski J. D., Lobanov V. S., Myslik J., Graf E., Carver T., Asel E., Springer B. A., Lane P. and Salemme F. R. (2001). "High-density miniaturized thermal shift assays as a general strategy for drug discovery." J Biomol Screen **6**(6): 429-40.
- Paoli P., Giannoni E., Pescitelli R., Camici G., Manao G. and Ramponi G. (2001). Hydrogen peroxide triggers the formation of a disulfide dimer of muscle acylphosphatase and modifies some functional properties of the enzyme. J Biol Chem, United States. **276**: 41862-9.
- Papworth C., Bauer J. C., Braman J. and Wright D. A. (1996). "QuikChange site-directed mutagenesis." Strategies **9**: 3-4.
- Parks T. D., Leuther K. K., Howard E. D., Johnston S. A. and Dougherty W. G. (1994). "Release of proteins and peptides from fusion proteins using a recombinant plant virus proteinase." Anal Biochem **216**(2): 413-7.
- Patil A. G., K P. K., Mulimani V. H., Veeranagouda Y. and Lee K. (2010). "alpha-Galactosidase from *Bacillus megaterium* VHM1 and its application in removal of flatulence-causing factors from soymilk." J Microbiol Biotechnol **20**(11): 1546-54.
- Pedersen N. C., Ho E. W., Brown M. L. and Yamamoto J. K. (1987). "Isolation of a T-lymphotropic virus from domestic cats with an immunodeficiency-like syndrome." Science **235**(4790): 790-3.
- Perugino G., Trincone A., Rossi M. and Moracci M. (2004). "Oligosaccharide synthesis by glycosynthases." Trends Biotechnol **22**(1): 31-7.

- Poiesz B. J., Ruscetti F. W., Gazdar A. F., Bunn P. A., Minna J. D. and Gallo R. C. (1980). "Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma." Proc Natl Acad Sci U S A **77**(12): 7415-9.
- Porath J., Carlsson J., Olsson I. and Belfrage G. (1975). "Metal chelate affinity chromatography, a new approach to protein fractionation." Nature **258**(5536): 598-9.
- Prakash K., Ranganathan P. N., Mettus R., Reddy P., Srinivasan A. and Plotkin S. (1992). "Generation of deletion mutants of simian immunodeficiency virus incapable of proviral integration." J Virol **66**(1): 167-71.
- Prasher J. M., Elenitoba-Johnson K. S. and Kelley L. L. (2001). "Loss of p53 tumor suppressor function is required for in vivo progression of Friend erythroleukemia." Oncogene **20**(23): 2946-55.
- Primrose S. B. and Twyman R. (2006). Principles of Gene Manipulation and Genomics, Wiley-Blackwell.
- Putnam C. D., Hammel M., Hura G. L. and Tainer J. A. (2007). X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. Q Rev Biophys. England. **40**: 191-285.

[R]

- Raines R. T., McCormick M., Van Oosbree T. R. and Mierendorf R. C. (2000). "The S.Tag fusion system for protein purification." Methods Enzymol **326**: 362-76.
- Rauter A. P. and Lindhorst T. (2011). Carbohydrate Chemistry: Chemical and Biological Approaches, Royal Society of Chemistry.
- Ren G., Gao K., Bushman F. D. and Yeager M. (2007). "Single-particle image reconstruction of a tetramer of HIV integrase bound to DNA." J Mol Biol **366**(1): 286-94.
- Rety S., Reaeabkova L., Dubanchet B., Silhan J., Legrand P. and Lewit-Bentley A. (2010). "Structural studies of the catalytic core of the primate foamy virus (PFV-1) integrase." Acta Crystallogr Sect F Struct Biol Cryst Commun **66**(Pt 8): 881-6.
- Reynolds J. H. (1974). "An immobilized alpha-galactosidase continuous flow reactor." Biotechnol Bioeng **16**(1): 135-47.
- Rhodes D. I., Peat T. S., Vandegraaff N., Jeevarajah D., Le G., Jones E. D., Smith J. A., Coates J. A., Winfield L. J., Thienthong N., Newman J., Lucent D., Ryan J. H., Savage G. P., Francis C. L. and Deadman J. J. (2011). "Structural basis for a new mechanism of inhibition of HIV-1 integrase identified by fragment screening and structure-based design." Antivir Chem Chemother **21**(4): 155-68.
- Rice P. A. and Baker T. A. (2001). "Comparative architecture of transposase and integrase complexes." Nat Struct Biol **8**(4): 302-7.
- Rossmann G. M. and Blow D. M. (1962). "The detection of sub-units within the crystallographic asymmetric unit." Acta Crystallographica **15**: 24-31.
- Rous P. (1911). "A sarcoma of the fowl transmissible by an agent separable from the tumor cells." J Exp Med **13**(4): 397-411.
- Ruprecht K., Mayer J., Sauter M., Roemer K. and Mueller-Lantzsch N. (2008). "Endogenous retroviruses and cancer." Cell Mol Life Sci **65**(21): 3366-82.
- Russell R. J., Gerike U., Danson M. J., Hough D. W. and Taylor G. L. (1998). "Structural adaptations of the cold-active citrate synthase from an Antarctic bacterium." Structure **6**(3): 351-61.

[S]

- Sanger F., Nicklen S. and Coulson A. R. (1977). "DNA sequencing with chain-terminating inhibitors." Proc Natl Acad Sci U S A **74**(12): 5463-7.
- Sassenfeld H. M. and Brewer S. J. (1984). "A Polypeptide Fusion Designed for the Purification of Recombinant Proteins." Nature Biotechnology **2**: 76-81.

- Sastri J. and Campbell E. M. (2011). "Recent insights into the mechanism and consequences of TRIM5 α retroviral restriction." *AIDS Res Hum Retroviruses* **27**(3): 231-8.
- Schwartzberg P., Colicelli J. and Goff S. P. (1984). "Construction and analysis of deletion mutations in the pol gene of Moloney murine leukemia virus: a new viral function required for productive infection." *Cell* **37**(3): 1043-52.
- Semenova E. A., Marchand C. and Pommier Y. (2008). "HIV-1 integrase inhibitors: update and perspectives." *Adv Pharmacol* **56**: 199-228.
- Shi K., Pandey K. K., Bera S., Vora A. C., Grandgenett D. P. and Aihara H. (2013). A possible role for the asymmetric C-terminal domain dimer of Rous sarcoma virus integrase in viral DNA binding. *PLoS One*. United States. **8**: e56892.
- Shuman H. A. and Silhavy T. J. (2003). "The art and design of genetic screens: *Escherichia coli*." *Nature Reviews Genetics* **4**: 419-431.
- Sigurdsson B. (1954). "Observations on three slow infections of sheep. Maedi, paratuberculosis, rida, a slow encephalitis of sheep with general remarks on infections which develop slowly and some of their special characteristics." *Br. Vet. J.* **110**: 255-270.
- Sigurdsson B., Palsson P. and Grimsson H. (1957). "Visna, a demyelinating transmissible disease of sheep." *J Neuropathol Exp Neurol* **16**(3): 389-403.
- Sloan R. D. and Wainberg M. A. (2011). "The role of unintegrated DNA in HIV infection." *Retrovirology* **8**: 52.
- Sloan R. D. and Wainberg M. A. (2011). The role of unintegrated DNA in HIV infection. *Retrovirology*. England. **8**: 52.
- Smith C. (2007). "Cloning and mutagenesis: tinkering with the order of things." *Nature Methods* **4**: 455-461.
- Spangenberg P., Andre C., Dion M., Rabiller C. and Mattes R. (2000). "Comparative study of new alpha-galactosidases in transglycosylation reactions." *Carbohydr Res* **329**(1): 65-73.
- Stehelin D., Varmus H. E., Bishop J. M. and Vogt P. K. (1976). "DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA." *Nature* **260**(5547): 170-3.
- Stick R. V. and Williams S. (2008). *Carbohydrates: The Essential Molecules of Life: The Essential Molecules of Life*, Cold Spring Harbor Laboratory Press.
- Strebel K., Luban J. and Jeang K. T. (2009). Human cellular restriction factors that target HIV-1 replication. *BMC Med*. England. **7**: 48.
- Strobel G. A. (1974). "Phytotoxins Produced by Plant Parasites." *Annual Review of Plant Physiology* **25**: 541-566.
- Struvay C. and Feller G. (2012). Optimization to low temperature activity in psychrophilic enzymes. *Int J Mol Sci*. Switzerland. **13**: 11643-65.
- Studier F. W. (1991). "Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system." *J Mol Biol* **219**(1): 37-44.
- Studier F. W. and Moffatt B. A. (1986). "Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes." *J Mol Biol* **189**(1): 113-30.
- Sugimoto J. and Schust D. J. (2009). "Review: human endogenous retroviruses and the placenta." *Reprod Sci* **16**(11): 1023-33.
- Suter-Crazzolara C. and Unsicker K. (1995). "Improved expression of toxic proteins in *E. coli*." *Biotechniques* **19**(2): 202-4.
- Suzuki H., Osawa Y., Oota H. and Yoshida H. (1969). "Studies on the decomposition of raffinose by β -galactosidase of mold." *Agricultural and biological chemistry* **33**: 501-513.
- Suzuki Y. and Craigie R. (2007). The road to chromatin - nuclear entry of retroviruses. *Nat Rev Microbiol*. England. **5**: 187-96.
- Svergun D. I. (1992). "Determination of the regularization parameter in indirect-transform methods using perceptual criteria." *Journal of Applied Crystallography* **25**: 495-503.
- Swanstrom R. and Coffin J. (2012). "HIV-1 pathogenesis: the virus." *Cold Spring Harb Perspect Med* **2**(12): a007443.

Szmigielski S. (1966). "An improved method for histochemical demonstration of N-glucuronidase and alpha-galactosidase activity in bone marrow and blood cells." Journal of Laboratory and Clinical Medicine **67**: 709.

[T]

Taylor R. G., Walker D. C. and McInnes R. R. (1993). "*E. coli* host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing." Nucleic Acids Res **21**(7): 1677-8.

Temin H. M. and Mizutani S. (1970). "RNA-dependent DNA polymerase in virions of Rous sarcoma virus." Nature **226**(5252): 1211-3.

Terwisscha van Scheltinga A. C., Armand S., Kalk K. H., Isogai A., Henrissat B. and Dijkstra B. W. (1995). "Stereochemistry of chitin hydrolysis by a plant chitinase/lysozyme and X-ray structure of a complex with allosamidin: evidence for substrate assisted catalysis." Biochemistry **34**(48): 15619-23.

Toney M. D., Hohenester E., Cowan S. W. and Jansonius J. N. (1993). "Dialkylglycine decarboxylase structure: bifunctional active site and alkali metal sites." Science **261**(5122): 756-9.

Tsukahara T., Agawa H., Matsumoto S., Matsuda M., Ueno S., Yamashita Y., Yamada K., Tanaka N., Kojima K. and Takeshita T. (2006). "Murine leukemia virus vector integration favors promoter regions and regional hot spots in a human T-cell line." Biochem Biophys Res Commun **345**(3): 1099-107.

Tsurutani N., Kubo M., Maeda Y., Ohashi T., Yamamoto N., Kannagi M. and Masuda T. (2000). "Identification of critical amino acid residues in human immunodeficiency virus type 1 IN required for efficient proviral DNA formation at steps prior to integration in dividing and nondividing cells." J Virol **74**(10): 4795-806.

Turelli P., Doucas V., Craig E., Mangeat B., Klages N., Evans R., Kalpana G. and Trono D. (2001). "Cytoplasmic recruitment of INI1 and PML on incoming HIV preintegration complexes: interference with early steps of viral replication." Mol Cell **7**(6): 1245-54.

[V]

Vagin A. and Teplyakov A. (2010). Molecular replacement with MOLREP. Acta Crystallogr D Biol Crystallogr. England. **66**: 22-5.

Valkov E., Gupta S. S., Hare S., Helander A., Roversi P., McClure M. and Cherepanov P. (2009). "Functional and structural characterization of the integrase from the prototype foamy virus." Nucleic Acids Res **37**(1): 243-55.

Valle-Casuso J. C., Di Nunzio F., Yang Y., Reszka N., Lienlaf M., Arhel N., Perez P., Brass A. L. and Diaz-Griffero F. (2012). TNPO3 is required for HIV-1 replication after nuclear import but prior to integration and binds the HIV-1 core. J Virol. United States. **86**: 5931-6.

Vallée H. and Carré H. (1904). "Sur la nature infectieuse de l'anémie du cheval." C.R. Hebd. Seances Acad. Sci. Ser. D Sci. Nat. **139**: 331-333.

Van de Weert M. and Stella L. (2011). "Fluorescence quenching and ligand binding: A critical discussion of a popular methodology." Journal of Molecular Structure **998**(1-3): 144-150.

Viana P. A., de Rezende S. T., Marques V. M., Trevizano L. M., Passos F. M., Oliveira M. G., Bemquerer M. P., Oliveira J. S. and Guimaraes V. M. (2006). "Extracellular alpha-galactosidase from *Debaryomyces hansenii* UFV-1 and its use in the hydrolysis of raffinose oligosaccharides." J Agric Food Chem **54**(6): 2385-91.

Vieille C., Burdette D. S. and Zeikus J. G. (1996). "Thermostables." Biotechnol Annu Rev **2**: 1-83.

Vieille C. and Zeikus G. J. (2001). "Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability." Microbiol Mol Biol Rev **65**(1): 1-43.

- Vink C., Lutzke R. A. and Plasterk R. H. (1994). "Formation of a stable complex between the human immunodeficiency virus integrase protein and viral DNA." Nucleic Acids Res **22**(20): 4103-10.
- Violot S., Hong S. S., Rakotoibe D., Petit C., Gay B., Moreau K., Billaud G., Priet S., Sire J., Schwartz O., Mouscadet J. F. and Boulanger P. (2003). "The human polycomb group EED protein interacts with the integrase of human immunodeficiency virus type 1." J Virol **77**(23): 12507-22.
- Vocadlo D. J. and Davies G. J. (2008). "Mechanistic insights into glycosidase chemistry." Curr Opin Chem Biol **12**(5): 539-55.
- Vogelstein B. and Gillespie D. (1979). "Preparative and analytical purification of DNA from agarose." Proc Natl Acad Sci U S A **76**(2): 615-9.
- Voisset C. and Andrawiss M. (2000). "Retroviruses at a glance." Genome Biology **1**(3): 4015.1-4015.4.
- Vuong T. V. and Wilson D. B. (2010). "Glycoside hydrolases: catalytic base/nucleophile diversity." Biotechnol Bioeng **107**(2): 195-205.

[W]

- Wang J. Y., Ling H., Yang W. and Craigie R. (2001). "Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein." EMBO J **20**(24): 7333-43.
- Warrilow D., Tachedjian G. and Harrich D. (2009). "Maturation of the HIV reverse transcription complex: putting the jigsaw together." Rev Med Virol **19**(6): 324-37.
- Webb E. C. (1992). Enzyme Nomenclature 1992: Recommendations of the NCIUBMB on the Nomenclature and Classification of Enzymes, Academic Press.
- Weiss B. and Richardson C. C. (1967). "Enzymatic breakage and joining of deoxyribonucleic acid, I. Repair of single-strand breaks in DNA by an enzyme system from *Escherichia coli* infected with T4 bacteriophage." Proc Natl Acad Sci U S A **57**(4): 1021-8.
- Weiss R. A. (2006). "The discovery of endogenous retroviruses." Retrovirology **3**: 67.
- Wielens J., Crosby I. T. and Chalmers D. K. (2005). "A three-dimensional model of the human immunodeficiency virus type 1 integration complex." J Comput Aided Mol Des **19**(5): 301-17.
- Wielens J., Headey S. J., Jeevarajah D., Rhodes D. I., Deadman J., Chalmers D. K., Scanlon M. J. and Parker M. W. (2010). Crystal structure of the HIV-1 integrase core domain in complex with sucrose reveals details of an allosteric inhibitory binding site. FEBS Lett. Netherlands, 2010 Federation of European Biochemical Societies. Published by Elsevier B.V. **584**: 1455-62.
- Wilcox W. R., Banikazemi M., Guffon N., Waldek S., Lee P., Linthorst G. E., Desnick R. J., Germain D. P. and International Fabry Disease Study G. (2004). "Long-term safety and efficacy of enzyme replacement therapy for Fabry disease." Am J Hum Genet **75**(1): 65-74.
- Wlodawer A., Hodgson K. O. and Shooter E. M. (1975). "Crystallization of nerve growth factor from mouse submaxillary glands." Proc Natl Acad Sci U S A **72**(3): 777-9.

[Y]

- Yamashita M. and Emerman M. (2006). "Retroviral infection of non-dividing cells: old and new perspectives." Virology **344**(1): 88-93.
- Yang F. and Roth M. J. (2001). "Assembly and catalysis of concerted two-end integration events by Moloney murine leukemia virus integrase." J Virol **75**(20): 9561-70.
- Yang Z. N., Mueser T. C., Bushman F. D. and Hyde C. C. (2000). "Crystal structure of an active two-domain derivative of Rous sarcoma virus integrase." J Mol Biol **296**(2): 535-48.

- Yao X., Fang S., Qiao W., Geng Y. and Shen Y. (2010). "Crystal structures of catalytic core domain of BIV integrase: implications for the interaction between integrase and target DNA." Protein Cell **1**(4): 363-70.
- Yoshimitsu M., Higuchi K., Fan X., Takao S., Medin J. A., Tei C. and Takenaka T. (2011). "Sequencing and characterization of the porcine alpha-galactosidase A gene: towards the generation of a porcine model for Fabry disease." Mol Biol Rep **38**(5): 3145-52.

[Z]

- Zeikus J. G., Vieille C. and Savchenko A. (1998). "Thermostzymes: biotechnology and structure-function relationships." Extremophiles **2**(3): 179-83.
- Zennou V., Petit C., Guetard D., Nerhbass U., Montagnier L. and Charneau P. (2000). HIV-1 genome nuclear import is mediated by a central DNA flap. Cell. United States. **101**: 173-85.
- Zhang R. and Monsma F. (2010). "Fluorescence-based thermal shift assays." Curr Opin Drug Discov Devel **13**(4): 389-402.
- Zhao H., Lu L., Xiao M., Wang Q., Lu Y., Liu C., Wang P., Kumagai H. and Yamamoto K. (2008). "Cloning and characterization of a novel alpha-galactosidase from *Bifidobacterium breve* 203 capable of synthesizing Gal-alpha-1,4 linkage." FEMS Microbiol Lett **285**(2): 278-83.
- Zhao L. and Chmielewski J. (2013). "Inhibition of HIV-1 integrase dimerization and activity with crosslinked interfacial peptides." Bioorg Med Chem **21**(14): 4041-4.
- Zheng R., Ghirlando R., Lee M. S., Mizuuchi K., Krause M. and Craigie R. (2000). "Barrier-to-autointegration factor (BAF) bridges DNA in a discrete, higher-order nucleoprotein complex." Proc Natl Acad Sci U S A **97**(16): 8997-9002.
- Zheng R., Jenkins T. M. and Craigie R. (1996). "Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity." Proc Natl Acad Sci U S A **93**(24): 13659-64.
- Zhu A., Leng L., Monahan C., Zhang Z., Hurst R., Lenny L. and Goldstein J. (1996). "Characterization of recombinant alpha-galactosidase for use in seroconversion from blood group B to O of human erythrocytes." Arch Biochem Biophys **327**(2): 324-9.
- Zhu K., Dobard C. and Chow S. A. (2004). "Requirement for integrase during reverse transcription of human immunodeficiency virus type 1 and the effect of cysteine mutations of integrase on its interactions with reverse transcriptase." J Virol **78**(10): 5045-55.